

# California against cronyism

The scale of funds set to be spent by the state on stem-cell research necessitates strong governance.

**T**he California Institute for Regenerative Medicine (CIRM) has been making up for lost time. It didn't award a single grant until more than a year after California taxpayers voted it into existence, because lawsuits were holding up its funding. Now the agency is about to award another round of grants, which will bring its total spending to just over \$520 million. The upcoming awards will fund a network of new and renovated buildings across the state — a striking achievement, all the more remarkable for coming in the same year that Governor Arnold Schwarzenegger asked the state's public universities to cut the rest of their budgets across the board.

So much for the CIRM's good news. Now for the growing pains. Several episodes over the past year have highlighted an inherent problem with the CIRM's structure: the board that distributes its funding is stacked with representatives from the universities that benefit most from those disbursements. The CIRM has enacted rules to try to limit the conflicts of interest posed by this arrangement. They don't go far enough. At one meeting in January, for instance, CIRM board members from institutions that had applied for a facilities grant voted to deny one of these grants to an institution that has no representatives on the CIRM board (see page 18).

Another potential conflict is sure to become a bigger issue if, as the CIRM has promised, its discoveries yield commercializable results. Some in the California state legislature, such as Senator Sheila Kuehl (Democrat), want to ensure that treatments resulting from the CIRM can be afforded by all taxpayers footing the bill. But the CIRM must also court biotechnology companies if its discoveries are to make it to the clinic, and these companies want maximum control over their pricing options.

The CIRM requires companies to which it gives money to offer

any resulting drugs “at a price as provided in the California Discount Prescription Drug Program” to people eligible for that programme. The CIRM also requires academic institutions to license their CIRM-funded inventions “only to persons that agree to have a plan in place at the time of commercialization to provide access to resultant therapies and diagnostics for uninsured California patients”.

But these relatively vague requirements lack teeth, a fact that Senators Kuehl and George Runner (Republican) are attempting to address with legislation requiring CIRM grantees and licencees to “sell drugs at a price that does not exceed any benchmark price in the California Discount Prescription Drug Program”, a measure aimed at ensuring affordability, and a sensible one at that. This proposal is a way to nudge the CIRM towards something it has already promised — guaranteed access for taxpayers to the results of the work they have funded. It also attempts to strengthen the voice of the public in setting the CIRM's path — a voice that, although institutionalized on the CIRM's board, has competed there with mixed success against academic and industrial interests.

This will be the CIRM's main procedural challenge. For the agency to succeed, patient advocates and other public representatives must fight the tendency of the academic institutions on the board to hoard dollars. As the patient advocates grow into their roles as full partners, and with help from well-intentioned lawmakers such as Kuehl, the CIRM must be coaxed into serving its most important constituency — the taxpayers of California. The roles themselves are not unusual in the world of governance, but here the stakes are exceptionally high. ■

**“The CIRM must be coaxed into serving its most important constituency — the taxpayers of California.”**

## A research menu

More spending on agricultural science is needed to help resolve the world's food crisis.

**T**his was not a sudden crisis. It may be only this spring that food prices have started sparking riots on the streets of Haiti and Egypt, not to mention rice rationing at Wal-Mart's cash-and-carry stores, but food prices have been rising since 2000. The rises accelerated in 2006, when global cereal stocks dropped to levels not seen since the early 1980s. And although the factors driving them are many and various, a good few of them look likely to persist for years to come.

Nor is the crisis unremittingly heinous. Higher food prices, other things being equal, mean higher farm incomes, and there are a lot of poor farmers in the world who could do with such a boost. But although this may suggest benefits for some in the future, the net

effect so far has been negative. An interim report released by the World Bank in April says that seven years' poverty reduction has been undone by the past two years of high staple-food prices.

The causes of these shortages are not easily undone, and some of them are things no one should want to undo. In China and India there is ever more — and utterly reasonable — demand for a third meal in the day and more meat in the diet. In Australia prolonged drought has had a severe effect on wheat production. High energy prices mean costly fertilizers and insecticides, not to mention making farm machinery more expensive to run. In the United States, more and more corn (maize) goes to making ethanol, raising the price of both corn and other cereals that can substitute for it.

There are various ways in which the fruits of scientific research might have helped ease the suffering that comes from this confluence of factors. But here, too, the harvest is not what it might have been. Public spending on basic agricultural research fell during the 1980s and 1990s in rich countries (see page 8). The proportion of US aid ploughed into agriculture wilted from 25% to 1%, bilateral farming aid from Europe

dropped by two-thirds and World Bank lending in the sector slipped from 30% to 8%. The reasons for this included the perceived success of Green Revolution technologies in Asia, and, indeed, some backlash against intensified farming among green groups. The downslide was most pronounced in sub-Saharan Africa, where the cutbacks were still severe even though there had been no Green Revolution comparable to that in Asia. A contributing factor to this decline from the 1990s on was Europe's attitude to genetically modified crops, which both chilled research in the area and reduced incentives for such technologies to be fielded in countries looking to European export markets.

One might assume that such cutbacks in research reflected poor results. Not so; the pay-offs to agricultural research are massive. The World Bank's *World Development Report 2008: Agriculture for Development* (<http://tinyurl.com/2ngyqd>) — the first of the annual reports to focus on agriculture for a quarter of a century, the bank noted with self-reproach — cites 700 published estimates of rates of return on investment in agricultural research, development and extension services in developing countries. It reports an average annual return of 43%.

Agriculture has poverty-busting powers beyond straightforward revenue increases. One reason for this is that poor people in poor countries who earn a little extra cash will spend it on basic local goods and services — agricultural growth spurs economic growth from the bottom up. A study of 42 developing countries covering the period from 1981 to 2003 found that growth in gross domestic product (GDP) that originated in agriculture increased spending by the poor two-and-a-half times more than does GDP growth in other sectors.

The past weeks have brought signs that global institutions and donors are beginning to bow, belatedly, to this logic. On 2 April the World Bank announced its intention to double agricultural lending to sub-Saharan Africa over the next year, and bank administrators say that a portion of the new money will go towards basic research.

Britain, the International Monetary Fund and the Bill & Melinda Gates Foundation are also opening their coffers. In the case of the Gates's money, much will be channelled through the Alliance for a Green Revolution in Africa led by Kofi Annan.

There are many useful directions for such development; higher yields, drought resistance and reduced requirements for inputs such as fertilizers and pesticides are all promising. But the more pressing problem for poor farmers is not the development of new technologies but access to those already there. Plenty of good agricultural science — such as locally adapted seed varieties and soil surveys — sits unused because it has not been delivered in a form adequately tailored to the end users and their limited means.

Resources need to go towards coordinating and strengthening local agricultural extension services as an integral part of revamping and reintegrating the research infrastructure. Agricultural research systems in sub-Saharan Africa are fragmented into almost 400 distinct agencies, eight times the number in the United States and four times the number in India.

Access is not just a matter of seeds and bloodlines and new agromonic know-how. Weather services that rich-world farmers rely on are extraordinarily hard to get hold of in poor countries. The first attempt to evaluate the impact of climate change on agriculture in Africa, for example, had to rely on climate-sensitivity studies carried out in the United States for lack of any data from the continent in question. Better weather and climate data would allow lenders, be they development banks or local sources of microfinance, to create insurance products for farmers of a sort that is almost completely unavailable to the world's poor. These pragmatic solutions should get a large slice of a rapidly expanded pie. ■

**"Agriculture has poverty-busting powers beyond revenue increases."**

## A place for everything

More researchers must record the latitude and longitude of their data.

**W**ho, what, where and when? Among the basic elements of scientific record-keeping, too often the 'where?' gets neglected. Now advances in satellite-positioning technology, online databases and geographical information systems offer opportunities to make good that neglect, and to add a much-needed spatial dimension to many types of biological research. Location data are essential for those modelling species' responses to climate change, or the spread of viruses, for example. Failure to include spatial information from the get-go may close off potentially highly productive routes to analysis — including those not yet foreseen. But those data are frequently inadequate or absent.

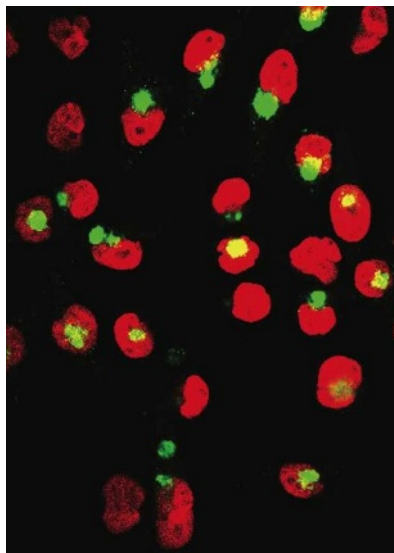
Many museums and herbaria are trying to make good this problem as best they can, geo-referencing their collections and putting them online. This frequently requires nightmarish work translating place names from various historical eras, languages and conventions into

latitudes and longitudes. Although this is a necessary evil in matters retrospective, going forward there is a much simpler and easier answer in the form of coordinates and a time-stamp taken from the Global Positioning System (GPS) at the point of capture, or any other specified point of relevance.

This technology means that there is now much less excuse for allowing spatial data to fall by the wayside simply because they are not relevant to the data collectors' project in hand. Not only are the data easily collected, they are easily stored too. GenBank, for example, introduced fields for latitude and longitude in the metadata attached to its nucleotide sequence records in 2005. But few yet contain such information.

Gene sequence and structure databases have flourished in part because journals require authors to submit published data to them. It is worth considering a similar requirement that all samples in a published study be registered, along with GPS coordinates, in online databases such as the Global Biodiversity Information Facility. At the same time, it would behove spatial scientists to articulate to the broader research community the potential of recording and making accessible spatial data in the appropriate formats — and the painlessness of the process. ■

# RESEARCH HIGHLIGHTS



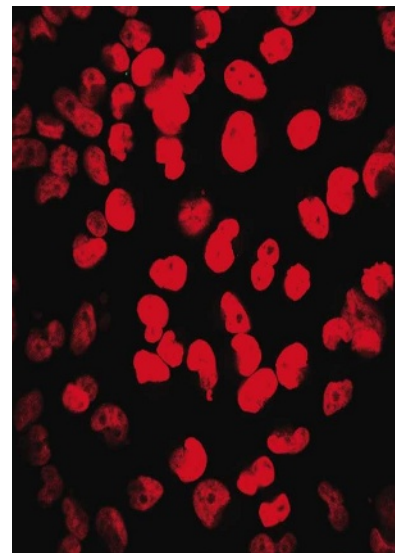
## Casting anchor

*Science* **320**, 520–523 (2008)

A potential drug target for Alzheimer's disease may be more tractable now that scientists in Germany have discovered how to tether compounds that block a crucial enzyme to the cellular compartment in which the enzyme does its deadly work.

This enzyme,  $\beta$ -secretase, cleaves the protein APP (labelled green, left) to make amyloid- $\beta$  peptide, which causes the plaques that contribute to neurodegeneration, in a compartment called the endosome. But soluble inhibitors do not reach this area when tested in cell cultures.

Kai Simons from the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden and his colleagues chemically modified one inhibitor by linking it to a fat-soluble sterol group. The resulting compound could anchor itself inside an endosome and efficiently block amyloid- $\beta$  generation there (pictured right). It also stopped the development of plaques when injected into a mouse model of Alzheimer's disease.



L. RAJENDRAN

## NEUROBIOLOGY

### Pecking order

*Neuron* **58**, 273–283 (2008)

Social hierarchy is a major determinant of health and mortality, yet how the brain processes group position has been shrouded in mystery. By monitoring blood flow in gamers' brains, Caroline Zink, Andreas Meyer-Lindenberg and their colleagues at the US National Institute of Mental Health in Bethesda, Maryland, have revealed distinct brain activity patterns that form in response to status cues.

The researchers set artificial hierarchies by assigning 72 volunteers a skill rank in a computer game that flagged onscreen opponents as superior or inferior players. But the opponents were really computers, and the games and ranks were rigged so that status was only perceived. One of the authors' discoveries is that brain regions associated with emotion or pain become busier when gamers are losing to inferior opponents.

## HUMAN BIOLOGY

### Boy appétit

*Proc. R. Soc. B.* doi:10.1098/rspb.2008.0105 (2008)

Pregnant women who munch through a lot of calories around the time they conceive may be more likely to bear boys than girls.

Fiona Mathews of the University of Exeter, UK, and her colleagues divided 740 British women into three groups according to what they remembered eating in early pregnancy and around the time they conceived. Only 45% of those in the group with the lowest energy intake carried boys, compared with 56% of women in the highest-calorie group.

These data support the Trivers–Willard hypothesis, say the authors, which predicts that females in good condition will have more male offspring.

## CHEMICAL BIOLOGY

### Dope hope

*Nature Chem. Biol.* doi:10.1038/nchembio.86 (2008)

Marijuana calms people down and relieves pain, but also has the negative effect of promoting memory loss. By blocking the breakdown of two naturally occurring compounds that bind to the same brain receptor as cannabis, scientists in California have found a way to boost the positive effects associated with cannabis use.

THC, the marijuana plant's psychoactive component, binds to a receptor called CB<sub>1</sub>, as do the brain's own cannabinoids 2-AG and

anandamide. John Casida of the University of California, Berkeley, and his colleagues used organophosphorus nerve agents to inhibit the enzymes that catalyse the destruction of these two compounds. Adding one particular nerve agent caused a more than tenfold increase in the levels of these chemicals in the brain. The finding could help pharmacologists design new drugs that relieve pain.

## ASTRONOMY

### Galactic beginnings

*Astronom. J.* **135**, 1968–1981 (2008)

It is the ultimate chicken and egg problem: Which came first, galaxies' stars or the black holes at their centres? David Alexander from Durham University, UK, and his collaborators are putting their money on the stars.

The group surveyed six young galaxies more than 10 billion light years away and found that their black holes were between sixty million and one hundred million times more massive than our Sun.

That may be huge, but given the number of stars in each galaxy and current models of how galaxies grow, the findings suggest that these black holes appeared after the stars started multiplying. The stars probably helped the black holes grow by feeding them gases leftover from the stars' own birth.

## ANIMAL BEHAVIOUR

### A reassuring tune

*Curr. Biol.* **18**, 576–579 (2008)

Zoologists have debated whether birds that look out for predators while the rest of their flock feeds are behaving selfishly or cooperatively. But few have asked why these



©2008 Nature Publishing Group

PIXLAND/PHOTOLIBRARY.COM



winged watchmen draw attention to themselves by making particular calls.

Andrew Radford at the University of Bristol, UK, and his colleagues observed groups of pied babblers (*Turdoides bicolor*; pictured right) in the Kalahari in Africa, and recorded the sentinels' sounds. Playing the calls back, they found that the group ate more when a watchman sang; the flock also spread out more and looked up less often.

Because the foraging gains of other birds seem to explain the behaviour, the authors argue that it is an example of cooperation. The close kinship between guard and guarded suggests that sentinel duty is a means by which individuals increase their genetic contribution to future generations.



reversibly with temperature.

The same process could cause an electrostatic attraction between icy grains exposed to cosmic rays in space, promoting their aggregation into proto-planetary lumps.

## MOLECULAR PHYSICS

### Slicing the ice

*J. Phys. Chem. B* doi:10.1021/jp073870c (2008)

Ice can generate a voltage when heated or cooled, say Richard Bell at Pennsylvania State University in Altoona, James Cowin of Pacific Northwest National Laboratory in Richland, Washington, and their co-workers. The trick is to deposit ions at  $-113^{\circ}\text{C}$  onto a 1-micrometre-thick ice film, thereby charging the surface of the film and rotating its water molecules, which become locked in position when the film is later cooled.

Normally, water molecules are randomly oriented in ice. Each one has an asymmetrical distribution of positive and negative charge that could create a voltage were many water molecules to become frozen in alignment. Cowin's team's method achieves a voltage as large as those generated by commercial materials and that varies steeply and

## VIROLOGY

### Rubbish imitation

*Science* **320**, 531–535 (2008)

The vaccinia virus had a dirty secret.

Researchers in Switzerland have revealed that it dupes cells into taking it in by mimicking the detritus that many types of cell would normally mop up.

The work is important because vaccinia typifies pox viruses such as smallpox. Jason Mercer and Ari Helenius of ETH Zurich watched fluorescently labelled virus particles trigger the membranes of their target cells to develop a spherical bulge called a 'bleb'. Blebbing proved crucial for infection, and a fat molecule called phosphatidylserine in the viral membrane proved crucial for virus-induced blebbing. Because the process normally deals with 'rubbish' in healthy cells, it flies under the radar of the immune system and thus enables pox viruses to evade detection as they spread between cells.

## ATMOSPHERIC SCIENCE

### Jetting away

*Geophys. Res. Lett.* **35**, L08803 (2008)

Three bands of fast-moving wind — known as jet streams — have shifted position in recent decades, according to a new analysis of weather data that were collected between 1979 and 2001.

Shifts in the jet streams are expected in a changing climate, and the work by Cristina Archer and Ken Caldeira, of the Carnegie Institution of Washington in Stanford, California, looked at observational records for any

such effects. They report that the northern jet stream weakened, and that one of the southern jet streams weakened whereas the other strengthened. All three jet streams moved higher in altitude and closer to the poles, by about 19 kilometres per decade.

## NEUROSCIENCE

### Algal vision

*Nature Neurosci.* doi:10.1038/nn2117 (2008)

Blind mice that had lost all the light receptors in their retinæ have had their vision partly restored thanks to a light-sensitive ion channel from a green alga, *Chlamydomonas reinhardtii*.

Botond Roska of the Friedrich Miescher Institute for Biomedical Research in Basel, Switzerland, and his colleagues inserted the gene for this ion channel into the DNA of specific mouse retinal cells, the ON bipolar cells. Formerly blind mice with these cells gained rudimentary sight because light opened algal ion channels in the modified retinal neurons, allowing positively charged ions to flood in and a signal to be sent to the visual region of the brain. The mice fled from strong light and responded to moving grid patterns.

## JOURNAL CLUB

**Kristi Anseth**  
University of Colorado, Boulder

**A biologist looks to 'click chemistry' for better three-dimensional tissue models.**

A hot topic in organic chemistry is the development of ways to neatly home in on a particular chemical group and cause a reaction to proceed extremely efficiently under mild conditions. Such highly optimized reactions have been grouped under the term 'click chemistry'. A commonly

cited example involves functional groups called azides and alkynes, which react to form triazoles with the aid of a copper catalyst.

Click chemistry has all sorts of uses, although few are in biology because the technique relies on toxic metal catalysts. However, Carolyn Bertozzi and her colleagues at the University of California, Berkeley, and the nearby Lawrence Berkeley National Laboratory recently demonstrated copper-free click chemistry in a living system (J. M. Baskin *et al. Proc. Natl Acad. Sci. USA* **104**, 16793–16797; 2007). These authors selectively — and

rapidly — labelled cell-surface polysaccharides with triazole bound to a fluorescent probe. The technique allows real-time imaging of cell surface molecules that are otherwise impossible to achieve.

This research throws open the door for a host of new applications for click chemistry. As a tissue engineer, I am particularly excited about exploiting it to make better gels for three-dimensional cell culture.

Physiological processes are routinely guided by interactions between cells and their tissue environment. Thus, a major hurdle

in tissue regeneration is knowing which biochemical signals must be recapitulated in cell culture, and how to present them at the appropriate time and place. Copper-free click chemistry could allow scientists to synthesize materials that deliver these signals at times that are governed by the physiological conditions in which the material resides. Next on my wish list is the ability to control the spatial organization of these reactions.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

# Doctors accused of doing illegal stem-cell trials

An apparently successful stem-cell treatment for urinary incontinence is now being questioned after it has emerged that clinical trials for the therapy may have been done illegally.

Patients who received the stem-cell procedure in operations at a hospital in Innsbruck, Austria, are now taking legal action against the hospital's management, in a scandal that goes right to the heart of the country's health ministry.

Next week, a civil court will hear the case of 70-year-old Dieter Bollmann from Berlin, who paid more than €11,000 (US\$17,000) for the procedure in 2005. He claims that urologists, led by Hannes Strasser, at the University Hospital Innsbruck told him that the treatment had a very high success rate, and implied that clinical trials had been completed. The consent form that Bollmann signed, and which *Nature* has seen, does not include the word 'experimental'. The operation did not improve his symptoms.

Bollmann is suing TILAK, the administrative body that runs the university hospital, for €4,000 damages, in a suit he originally filed in 2006. Around that time, TILAK told Strasser and his team not to do any further operations outside of a formal clinical trial. A TILAK spokesman has told local newspapers that the team carried out a further 60 or so. In all, hundreds of patients have been treated outside

of clinical trials, it is alleged. A preliminary investigation by Austria's ministry of health has found that such patients are not insured for experimental procedures.

And the issue goes deeper. The independent ethics committee that approves and oversees clinical trials at the university and hospital is now concerned that the clinical trial of the procedure's efficacy, carried out by Strasser's team and published in *The Lancet* last June, may have been illegal. The Ministry of Health is now investigating claims by the ethics committee that it was never contacted, as law requires, about the trial on 63 women who were recruited between 2002 and 2004 (H. Strasser *et al. Lancet* 369, 2179–2186; 2007). And some

members of the committee claim that documents may have been created to cover up omissions. Further probing reveals apparent irregularities that trace back to the ministry itself.

Strasser's involvement with the ethics committee began in 2001, when he applied to conduct phase I pilot studies on his technique, to establish its safety. His procedure involves culturing muscle stem cells and fibroblasts from a small sample of tissue taken from a patient's arm. The cultured cells, produced by Innovacell Biotechnology, a company co-founded and co-owned by Strasser, are then injected into the same patient's urinary sphincter, the muscle at the base of the bladder that

**"It is important to repeat the good results in independent sites."**



MEDICAL UNIV. INNSBRUCK

**Clinical trials led by Hannes Strasser have come under scrutiny by an Austrian ethics committee.**

controls urine flow, where the cells are intended to promote new muscle growth and so increase the power of the sphincter.

The committee initially denied approval because the submission lacked data on, for example, animal experiments, which are necessary before new medical interventions can be used in humans. A year later the scientists provided these data and obtained approval from the Ministry of Health's technical committee. The university ethics committee then approved two phase I studies, in 2002 and 2003, in a total of 25 patients. Eventually, just 21 were recruited.

It was not until March 2007 that the urologists applied formally to undertake a phase II clinical trial — to assess the efficacy of their technique — on 100 patients. They included the draft of their *Lancet* paper in the supporting data they provided to the ethics committee. This was how the committee learnt that a randomized, controlled trial of 63 women had already been conducted without its approval.

The trial published in *The Lancet* states that it is registered on controlled-trials.com, however the reference number provided leads not to a national trial database but to an approval for the production of therapeutic cells by Innovacell.

The authors of the *Lancet* paper say that their trial was approved by the Ministry of Health on 27 August 2002. It is illegal under European (and Austrian) law for a trial to get the go-ahead without independent ethics approval. "The last time that the government gave approval for experiments on humans



**Patients treated outside of clinical trials in the Urology Department (inset) of the Medical University of Innsbruck may not have been insured.**

nwn

## NEXT WEEK'S NEWS

Vote on which story you would like us to cover in more depth.

[www.nature.com/news](http://www.nature.com/news)

without the input of an independent ethics committee was during the Third Reich," says Andreas Scheil, a professor of law at the University of Innsbruck and a member of the ethics committee.

Last October, in response to the committee's inquiries, Hubert Hrabcik, the director of the Ministry of Health's office for public health, wrote a letter to the committee saying that the legal situation was ambiguous but there seemed to be no obligation to approach an independent ethics committee.

But shortly after Hrabcik sent his letter, the urologists surprised the committee by suddenly producing the first page of an application to the health ministry for a phase III study of 90–100 patients, with the same registration number as the 2001 application — and dated to the same month — as the phase I study that had been denied. The committee says that it never saw this document, and some members have questioned its authenticity.

Strasser and his department head Georg Bartsch, who is also a witness in the Bollmann case, are currently under a gagging order. But their lawyer Dietmar Czernick told *Nature* that they did not require ethics approval for operations done outside of clinical trials. These treatments began after the clinical trial had been completed, although not yet published, Czernick claims, and at that point they considered the procedure to be "well-established therapy".

Czernick claims that Strasser and Bartsch believed they didn't need ethics approval for the trial itself because the law was "complicated and unclear" on the point. "They had the approval of the federal Ministry of Health which did the job of an ethics committee", Czernick says.

Financial issues are muddying the situation. TILAK paid an estimated €1.2 million for Innovacell materials for 200 patients, for trials it now thinks may have been uninsured and unapproved. Normally the sponsor of a trial — in this case Innovacell — pays such costs itself.

No one is suggesting that the procedure is harmful to patients. But researchers carrying out the treatment at sites outside Austria have not had the high success rate reported by the Innsbruck team. "It looks tremendously good, but as always it is important to repeat the good results in independent sites," says Giacomo Novara, a urologist at the University of Padua, Italy.

*The Lancet* is publishing an 'expression of concern' in its 3 May issue about the confusing reference number in the registration document in the trial they published. "We will wait for the outcome of the investigations before deciding about further actions," says the journal's senior executive editor Sabine Kleinert.

Alison Abbott

## US ocean-research projects in dire economic straits

America's fleet of research ships is struggling in financial doldrums threatening marine projects around the globe. The number of vessels is shrinking, funding for new vessels is being sidetracked and the forecast is for even fewer ships and higher costs. Already, the annual number of research days at sea has been cut by 20%.

"When the cost of equipment is driving the agenda, that is a death knell for a field," says geophysicist Marcia McNutt of the Monterey Bay Aquarium Research Institute in Moss Landing, California, who chairs the US research fleet's governing council. "You end up doing what research you can do, instead of what research you should do."

The fleet is overseen by the University-National Oceanographic Laboratory System (UNOLS), a consortium of 61 research institutes. The ships are affiliated to various institutions and provide vital platforms for scientists investigating topics from ocean chemistry and circulation to sea-floor core extractions. Annually, the fleet now provides about 4,000 research days at sea, down from about 5,000 before 2000, UNOLS officials say.

The UNOLS receives US\$80 million a year for fleet operation directly from the federal budget. This year, the funds have been supplemented by money from a private foundation and a Saudi university. Research teams apply to funding agencies, such as the US National Science Foundation (NSF), for individual scientific cruises — ship time for a deep-sea cruise runs at about \$50,000 per day.

There are 23 ships currently in service, varying from regional vessels for near-coast studies to oceanic cruisers. In 2002, there were 27 ships. By 2017, 15 ships are

projected, with only 11 by 2025 — at which time the global vessels would be down to three from today's six.

It takes years to design and refit or build a ship. For more than four years, the NSF, which funds most ship projects, has been planning to build a \$123 million Alaska-region research ship to replace a decommissioned vessel. But there is no ship funding in the NSF budget that is now before Congress — and long-term funding is unknown.

The ship *RV Marcus Langseth*, which will be the first ship to carry out three-dimensional models of the undersea crust, offers an example of the challenges faced. The newly rebuilt *Langseth* undertook her first research cruise in February after long delays and vast cost-overruns, involving difficulties securing high-tech crew members who are also sought by oil and gas companies, and costly problems operating its seismic system for visualizing seafloor formations. NSF officials say an audit is underway on the *Langseth* project, where costs at a Nova Scotia shipyard were \$600,000 above the bid of \$4.4 million for refitting. A worldwide lack of available shipyard space means that costly delays are systemic in the shipbuilding industry.

The *JOIDES Resolution*, a core-drilling vessel, is another delayed ship. It was to have been rebuilt by the end of 2007, but now isn't expected to be ready until at least August. The Integrated Ocean Drilling Program, an international partnership of scientists and research institutions that operates the ship, says the Singapore shipyard doing the refit will consume all contingency funds in the \$32 million retrofit budget, funded by the NSF.

Rex Dalton

The costly *RV Marcus Langseth* set off on her first research cruise in February.



LAMONT-DOHERTY EARTH OBSERVATORY



# Food crisis spurs research spending

More than 20 United Nations development agencies joined the World Bank and the World Trade Organization this week in Bern, Switzerland, to discuss emergency humanitarian aid and other measures to combat the growing world food crisis. The World Food Programme says it needs an extra US\$755 million just to meet existing needs for food aid.

The acute crisis is mainly due to market causes, but it is throwing the spotlight on the underlying problem of chronic underinvestment in agricultural research, experts say.

It's a wake-up call, says Nienke Beintema, who heads the Agricultural Science & Technology Indicators (ASTI) initiative at the International Food Policy Research Institute in Wageningen in the Netherlands. "Agricultural research had fallen off the international agenda," she says. "The food crisis has brought into focus the need for agriculture for economic development — and that people need food."

"It is now back on the agenda, and hopefully will translate into more investment in agricultural research," she says. There are already signs that this may be the case. For example, last week the United Kingdom announced an \$800 million, five-year package for agricultural research in developing countries.

Such funding is badly needed, according to the International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD), a four-year, \$10-million project with a structure of some 400 experts, akin to the Intergovernmental Panel on Climate Change. In September it is due to deliver its final assessment on how science, together with farming and economic practices, can best help to feed the poor in developing countries. But a draft version of its report warns that investment in agricultural research and its application to farming in poor countries is decreasing "at a time when it is most needed".

Overall, public spending on research such as pest and disease control and high-yielding crop varieties is growing in developing countries. However, this growth is largely accounted for by just four countries: China, India, Brazil and South Africa, according to the report. In many other developing countries, home to hundreds of millions of people, the report finds that spending is "stagnating or slipping".

That trend is all the more worrying because developing countries are benefiting less and less from transfer



Emergency aid helps now, but agencies are looking at longer-term strategies to alleviate food shortages.

of the results of intensive agricultural research carried out by rich countries, says Frances Kimmins, a lead author of the IAASTD report. The rate of growth in investment in agricultural research by industrialized countries ground to a halt following rapid growth from the 1960s to the early 1980s after decades of food surpluses. And focus has shifted away from improving crop varieties towards food processing and other 'added-value' products, which are less relevant to farmers in developing countries.

In developed countries, most agricultural research is increasingly done by the private sector, which will not invest in crops and practices for which there are small or nonexistent markets, and where patents prevent the spread of technologies. Taken together, these trends represent a drying-up of past sources of technology transfer from wealthy to developing countries, and make a case for renewed public

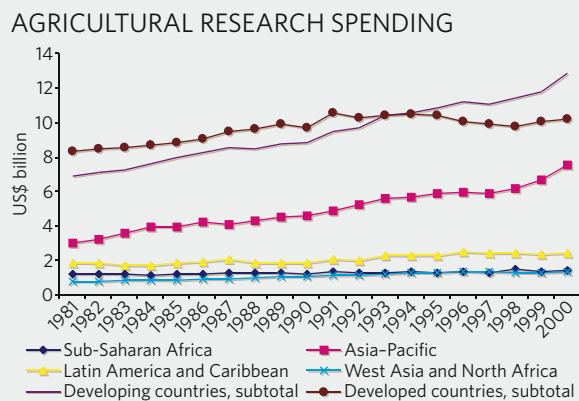
investment in developed nations, says Kimmins, who is also an agricultural expert at the UK-based development consultancy NR International.

With research spending high in China, India, Brazil and South Africa, these countries may take up the technology transfer gap left by rich countries, but it is too soon to tell. "Cooperation between these countries and the rest of the developing world is limited — it's only starting up," says Beintema. But the potential is there, she says, because they often share common agricultural climate zones.

Beintema's ASTI initiative is involved in a project with the Bill & Melinda Gates Foundation to explore the potential for greater agricultural research collaborations between developing nations. One of the first issues they will tackle is getting better data to guide science policy. "I've been surprised by how poor the data are," says Beintema, who is another lead author of the IAASTD report. She points out that the most recent statistics on agricultural research date from 2000 and are full of holes — and that included data for developed countries. A priority for the project with the Gates Foundation will be to get a better grasp of technology transfer and other issues by reinforcing the capacity of developing countries to collect the data needed, she says.

**Declan Butler**

See Editorial, page 1.



P. RAHMAN/AP

SOURCE: IAASTD REPORT



**HAVE YOUR SAY**

Comment on any of our news stories, online.

[www.nature.com/news](http://www.nature.com/news)

# Genetics bill cruises through Senate

The unanimous vote last week by the US Senate to outlaw discrimination against people on the basis of their genetic information is being celebrated by civil-rights groups, which have long campaigned for the safeguards. Personal-genomics companies are also cracking open the champagne — they have a lot to gain from the bill becoming law.

Once enacted, the Genetic Information Nondiscrimination Act (GINA) will forbid employers and health insurers from using people's genetic information against them in decisions on hiring, firing, promotion and insurance coverage and pricing. The House of Representatives should pass the legislation this week, after which it is expected to be signed into law by President George W. Bush.

"This will help the notion of personalized medicine move forward more quickly," says Linda Avey, co-founder of 23andMe, a personal-genetics company in Mountain View, California, that is trail-blazing a highly visible, and controversial, direct-to-consumer market. Avey says her employees erupted in cheers and applause when the bill's passage was announced at a staff meeting the next day. "We were very happy."

The bill is also likely to help other companies that trumpet the virtues of consumers' access to their own genetic data in a way that presumes it won't explode in their hands. "The customers of these new personal-genomics companies are able to download their genomes and share them electronically with others," notes Kathy

Hudson, director of the Genetics and Public Policy Center at Johns Hopkins University in Washington DC. "Until the passage of GINA, the sharing of that information actually put them at risk."

The bill bans US employers from collecting genetic information from their employees, and ensures that insurers can't request or require people to take genetic tests. Sanctions include government fines and lawsuits in federal courts. The House passed a similar bill a year ago, by 420 to 3. It was then sent to the Senate, where it was stalled by objections from Senator Thomas Coburn (Republican, Oklahoma).

The last of those objections was resolved last month by the insertion of wording preventing companies that insure their own employees from being punished twice under the law: once as an employer, and once as an insurer. But the new language also prevents an employee from suing their employer under the act if both the employer and the insurer are culpable in the same situation.

The wording is ambiguous enough, however, that it "will almost assuredly lead to litigation once it's passed into law," says Jeremy Gruber, the legal counsel for the National Workrights Institute, an employee advocacy group based in Princeton, New Jersey. It will therefore be the courts, he says, that "will clarify which situations might be objectionable enough" for an employer to be sued even if an insurance issue is involved.

The United States is not the first to implement such a law — countries including

Austria and France have laws forbidding genetic discrimination — but it has by far the largest private-insurance market.

The bill took months to get through the Senate and it still has its detractors. The Chamber of Commerce fought the bill on Capitol Hill, claiming that it would burden businesses with paperwork and expense, in part because it doesn't pre-empt a patchwork of existing state laws. "The bill also includes excessive damage provisions that will invite frivolous litigation," the Chamber continues to complain on its website.

But the bill's supporters argue that, rather than burdening the US employers who largely pay for that insurance, it will help them by easing health-care costs. "If we provide these protections, individuals will have the incentive to increasingly avail themselves of medical knowledge," says Senator Olympia Snowe (Republican, Maine), the leading sponsor of the Senate bill. "They may be able to take action as a result, preventing disease or premature death and also reducing the burden of high health-care costs."

For researchers, the law may prove a boon. The next generation of studies to identify gene culprits associated with complex diseases will involve tens of thousands of willing participants as cases and controls. "The success of those kinds of studies, I think, was significantly threatened by people's fears about genetic discrimination," says Hudson.

Meredith Wadman

## SNAPSHOT

### Rodent round-up

Zoologists have had the Pied Piper's knack of finding rodents this week. This greater white-toothed shrew (*Crocidura russula*) was discovered in Ireland after biologists were alerted to owl pellets containing large shrew skulls. The large rodent — well, larger than its native Irish cousins such as the pygmy shrew — probably arrived on the island as a stowaway on a ship from North Africa or continental Europe, where it is usually found. Several have now been found in Tipperary and Limerick.

Also reported this week, was the extraordinary find by biologists exploring the mossy mountain forests of Luzon island in the Philippines: a dwarf cloud rat. The species had not been seen since its 1896 discovery and naming by a British biologist. Sadly the reddish-brown rat was dead when spotted high in the canopy of Mount Pulag, the country's second-highest peak. "The cloud rats are one of the most spectacular cases of adaptive radiation by mammals anywhere in the world," says co-discoverer Lawrence Heaney of the Field Museum in Chicago, Illinois.



QUEEN'S UNIV. BELFAST/AFP PHOTO



# VIRUSES FOUND IN LUNG TUMOURS

Common viruses might contribute to lung cancer.  
[www.nature.com/news](http://www.nature.com/news)

GETTY

# If you go down to the woods today ...

## TEXAS

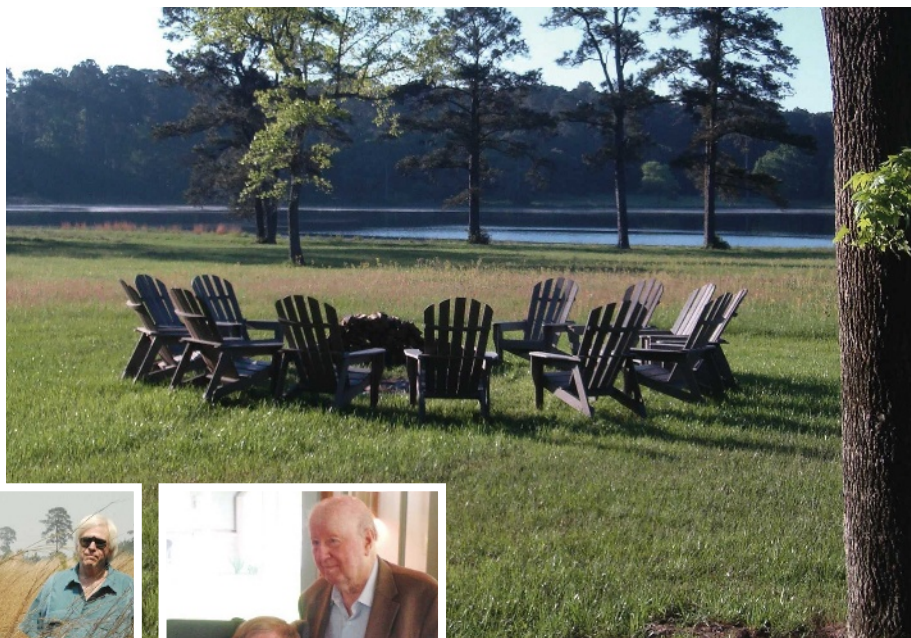
The Cook's Branch Conservancy in Montgomery County, Texas, is a large nature reserve of pine forest and wildflower meadows that is home to wild turkeys, bald eagles and a large population of the endangered red-cockaded woodpecker. For a fortnight in April, it was home to a rather different rare bird: the theoretical physicist.

A dozen of cosmology's brightest minds, including British astrophysicist Stephen Hawking, descended on the reserve to discuss the tricky problems of the early Universe. These physicists, most of whom are connected in some way to Hawking, either as collaborators or past graduate students at the University of Cambridge, UK, arrived for the invitation-only retreat, which, in its second year, has become one of the most exclusive — and pampered — workshops in physics.

The 23-square-kilometre property is owned by George Mitchell, an 88-year-old developer and oilman worth US\$3.2 billion. Late in life, Mitchell has cultivated a love for astrophysics, bestowing \$50 million on Texas A&M University in College Station. "I am trying to see how our top universities can have as much influence in high-level physics as, say, Caltech and the University of California, Berkeley, or Harvard or Yale," Mitchell explains. "And I am trying to see how we can get in on the act, because this state is big enough and wealthy enough to get it done."

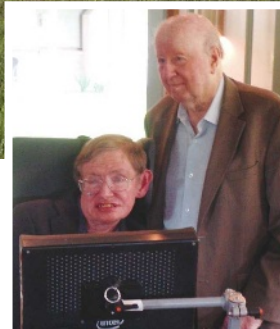
Mitchell is not the only billionaire donating to the theoretical physics cause. Mike Lazaridis, founder of Canada's Research in Motion, which makes the BlackBerry, gave \$100 million to kickstart the Perimeter Institute for Theoretical Physics in Waterloo, Canada. And James Simons, a mathematician who runs Renaissance Technologies, one of the world's largest hedge funds, in February gave \$60 million to Stony Brook University in New York to establish the Simons Center for Geometry and Physics.

But it somehow seems more personal with Mitchell, who wanted to study astronomy before deciding that petroleum engineering at Texas A&M would be more affordable (and lucrative). Not only has he given \$35 million to construct two new physics buildings, due for completion in 2009, he has also opened up the bucolic Cook's Branch estate — where the Mitchell family still gathers every year for Thanksgiving



E. HAND

E. HAND



Malcolm Perry, Gordon Kane, Stephen Hawking and George Mitchell (left to right) were among the elite physicists invited to the Cook's Branch Conservancy (above).

— as a luxurious retreat centre for collaborations involving Texas A&M physicists. And in April, the secluded lodges of Cook's Branch played elaborate host to Hawking and his colleagues. A Dutch chef from New York arrived to cook French cuisine for the week. One night, a classical guitar player serenaded the scientists as they ate a barbecued feast of wild pigs from the estate. And the wine flowed. Nick Warner, a string theorist from the University of Southern California in Los Angeles, says the nurturing environment is good for the mind. "A different exterior landscape can change the interior landscape," he says.

The physicists stayed up late around a fire, poking at each others' theories, and during the days tested new ideas in PowerPoint presentations. Andrew Strominger, of Harvard University in Cambridge, Massachusetts, gave a talk on 'chiral gravity in three dimensions', but he explained it in 2D, wheeling out two chalkboards and quickly camouflaging them in a jungle of algebra.

The physicists followed the formulae, computing in their heads. Strominger's talk

**"A different exterior landscape can change the interior landscape."**

was interrupted only by a cool breeze and the beep from Hawking's computer, a reminder of his only mode of communication. Hawking, who is incapacitated by amyotrophic lateral sclerosis, has a sensor, super-glued to a paperclip on the inside of his glasses, which registers a wrinkling of his cheek. Letter by letter, word by word, he composes sentences that are spoken by an old 1980s-era voice machine that he has kept because it has come to define him.

Texas A&M graduate students, invited out to Cook's Branch for the afternoon, rolled into the audience wearing baseball caps, shorts and sports sandals. They were duly impressed by the celebrities in attendance — but not too impressed, because, being in a physics department in the making, they aren't unaccustomed to seeing top physicists stop by College Station for a lecture or a guest appointment. The Mitchell money has helped Texas A&M attract two-dozen new physics faculty staff since 2001, and start an astronomy group from scratch.

"We're getting used to good people coming through," says graduate student Jim Ferguson, as he was served lobster bisque for lunch. ■

Eric Hand









Antarctica's climate may once have been akin to that now seen in southern Patagonia.

ANDRILL

## Sediment cores reveal Antarctica's warmer past

### VIENNA

A unique drilling project in the western Ross Sea has revealed that Antarctica had a much more eventful climate history than previously assumed. A new sediment core hints that the western part of the now-frozen continent went through prolonged ice-free phases — presumably offering a glimpse of where our warming world might be heading.

Researchers reported initial results from ANDRILL, a US\$30-million international drilling project, on 16 April at the assembly of the European Geosciences Union in Vienna. During the past two years, the team has extracted two cores, each containing some 1,200 metres of sediment, from the seabed below the vast Ross Ice Shelf, a floating extension of the West Antarctic Ice Sheet. Together, the cores provide an almost uninterrupted 17-million-year record of Antarctica's climatic past.

Palaeoclimatological records from ice cores, although more detailed and easier to interpret, cover only the past 800,000 years or so. Now, geologists say, Antarctica's history is laid out much more clearly.

"We have every page of the book," says David Harwood, an ANDRILL scientist at the University of Nebraska in Lincoln.

Sediments in the cores, along with microfossils such as pollen and spores, allow researchers to reconstruct sea temperatures and environmental conditions, such as the presence or absence of ice, over millions of years. The analytical work has only just begun, but early results indicate that, during warmer periods, the West Antarctic Ice Sheet and the Ross Ice Shelf shrank rapidly and substantially.

During a warm period some 3.5 million years ago, for instance, the ice sheet may have disap-

peared completely for around 200,000 years, raising sea levels globally by up to 10 metres.

For the first time, the ANDRILL cores show exactly how ice retreated rapidly and quickly in Antarctica. "That happened at a time when it was three to four degrees warmer than today, owing to atmospheric carbon dioxide concentrations, which we will very likely reach again soon," says Tim Naish, a project leader at the Institute of Geological and Nuclear Sciences in Lower Hutt, New Zealand.

In a poster session at the conference, David Pollard, an ice-sheet modeller at Pennsylvania State University in University Park, presented new simulations confirming that western Antarctica can lose almost all of its ice at temperatures just three to five degrees higher than today's. (During the past 1 million years, Antarctica's ice sheets have expanded.)

The first ANDRILL core covered the palaeoclimatic history of the past 7 million years. The second core, drilled between October and December last year, extends the record back to 17 million years ago. It reveals that during a period about 16 million years ago, Antarctica had a climate similar to that found today on the South Island of New Zealand and in southern Patagonia.

"It appears there were rivers, rain and trees in abundance," says Harwood. "That's really the biggest surprise to me so far."

What is still unclear is exactly how, when and where Antarctica's ice sheets started forming during the greenhouse-icehouse climate transition 35 million to 40 million years ago. To find out, the ANDRILL team hopes to resume drilling in 2012 at a new site on the Ross Ice Shelf.

Quirin Schiermeier

## International consortium to tackle cancer genomes

Research organizations from ten countries are forming an umbrella group to generate a comprehensive catalogue of genetic mutations in up to 50 types of cancer over the next ten years.

The International Cancer Genome Consortium was announced on 29 April. It will be coordinated by a secretariat at the Ontario Institute for Cancer Research in Toronto and will include the UK Wellcome Trust, France's National Cancer Institute, Japan's RIKEN and National Cancer Center and the US National Institutes of Health (NIH). Each participant is expected to spend some US\$20 million analysing genomic change in at least one type or subtype of cancer in roughly 500 patient samples. The project aims to avoid duplication and waste by coordinating the cancer types studied and by establishing common standards of data collection and analysis.

The NIH contribution will be made by the ongoing Cancer Genome Atlas project, an ambitious effort in its pilot phase aiming to analyse genes from some 50 types of cancer. The consortium is inviting other founding countries to join by 1 September.



Researchers on *Polarstern* recover a mooring.

## Warming Antarctic waters begin to cool

Antarctica's deep ocean waters are getting colder after years of warming, say researchers who have just returned from a Southern Ocean voyage aboard the German research vessel *Polarstern*.

Samples from previous expeditions showed that water at a depth of 4,500 metres in the Weddell Sea warmed by a tenth of a degree Celsius between 1989 and 2005, although the warming trend may have begun earlier. The latest work, by researchers from the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, found that temperatures have cooled slightly since 2005, suggesting that more cold surface water is reaching the deep ocean, perhaps as a result of changes in sea-ice coverage and atmospheric conditions. The team plans to revisit the region during the summer of 2010–11.

## UK government slammed for underfunding research

An influential committee of UK politicians has accused the government of failing to provide an adequate level of funding to sustain the country's science research.

Although the overall science budget is set to increase by 17.5% by 2011, changes within the structure mean that the actual amount of funded research will decline, charges the report from the cross-party House of Commons science select committee. That's because the budget will also need to pay for new bodies, such as the Technology Strategy Board, as well as the full economic costs of research. "We are concerned that the government has failed to protect both the existing and planned research base," says the report.

The report also has harsh words for the Science and Technology Facilities Council, which distributes research funding for physics. It says that the formation of the

council in 2007 from the merger of two previous councils was "untimely and poorly conceived", and that there are weaknesses in the council's peer-review system and management. The committee concludes that "substantial and urgent changes" are needed.

## Creationists fail in bid to offer 'science' degrees

A religious group has had its application to offer Master of Science degrees rejected by Texas authorities.

The Institute for Creation Research — which backs a literal interpretation of the Bible, including the creation of Earth in six days — was seeking a certificate to grant online degrees in science education in Texas (see *Nature* 451, 1030; 2008). But the Texas Higher Education Coordinating Board voted unanimously last week not to grant the institute's request, following the recommendation of Raymund Paredes, the state's commissioner of higher education.

"Religious belief is not science," Paredes said in his recommendation. "Science and religious belief are surely reconcilable, but they are not the same thing."

The institute has 45 days to appeal or 180 days to reapply.

## Artist cleared over possession of bacteria

A US district court in Buffalo, New York, last week cleared Steven Kurtz, an art professor at the State University of New York at Buffalo, of criminal charges related to the possession of bacterial cultures.

Four years ago, Kurtz found his wife dead at their home from a heart attack, and police responding to his emergency call discovered lab equipment and cultures there. Kurtz used biological materials to create politically charged art on topics such as government policy on genetically modified crops.

Initially investigated on charges of bioterrorism (see *Nature* 429, 690; 2004), Kurtz was indicted in 2004 for mail and wire fraud after receiving bacterial samples through the post. The case inspired the 2007 film *Strange Culture*.

In February, Robert Ferrell, a researcher at the University of Pittsburgh in Pennsylvania who sent the samples to Kurtz, pleaded guilty to mailing an injurious article and was fined \$500.

### Correction

The News story 'Data show extent of sexism in physics' (*Nature* 452, 918; 2008) incorrectly stated the percentage of talks given by female physicists on the DZero collaboration. Between May 2007 and May 2008, women gave 15% of all talks by the group.

## A lift for Europe's satellites

Galileo, Europe's planned satellite navigation system, is closer to reality following the launch of the long-delayed second verification satellite, Galileo In-Orbit Validation Element (GIOVE-B). The satellite was lofted into space from Baikonur, Kazakhstan, on 26 April, carrying an ultra-accurate atomic clock identical to those that will be used on the 30 planned operational Galileo satellites, set to be fully deployed by the end of 2013.

During the mission, GIOVE-B will discover what the environment is like at the altitude chosen for Galileo, and will send back signals like those intended to be used by the operational Galileo system.

The launch came days after the European Parliament voted to provide the €3.4 billion (US\$5.3 billion) needed to fund Galileo after a number of commercial partners pulled out of the project last year.



# Demonstrably wrong

Public-private demonstration projects are a good way to test technology in the field. But the driving force is often more political than scientific, argues **David Goldston**.

**T**he US government will spend about \$2.4 billion this fiscal year on applied research into energy efficiency and renewable sources, coal, oil, gas and nuclear energy. Much of the work will be done by researchers at national laboratories, but the most visible aspect of the US energy programme is demonstration projects — efforts jointly funded by the government and industry to test whether and how new technologies actually work in the field.

With their relatively large budgets, tangible facilities and corporate recipients, demo projects are ballyhooed by politicians and become rallying points for their communities, which benefit from the cash and jobs they bring. Demos tend to be the most politically secure yet substantively questionable activities in energy research and development.

There's perhaps no better illustration of the pitfalls of demonstration projects than the current brouhaha over FutureGen, the Bush administration's marquee demonstration project. FutureGen was proposed by the Department of Energy early in 2003 as a way to advance the president's climate technology and hydrogen initiatives. The plan called for building a new, state-of-the-art coal-burning electric plant that could capture its carbon dioxide emissions, which would be stored underground. The plant would also produce hydrogen gas and use it to generate additional electricity.

But this January, the Department of Energy abruptly announced that it was "restructuring" the project — actually cancelling it and replacing it with a programme to finance carbon capture and storage projects at privately built coal plants. The hydrogen aspect of the programme was dropped entirely. In explanation, the agency argued that the original project was becoming unsustainably costly and that the private sector was building more advanced coal plants on its own than had been expected five years ago.

Those are good reasons for cancelling FutureGen, but what really seems to have happened is that the coal industry for once overplayed its hand. In theory, industry is supposed to bear a sizable portion of the risk in demo projects — after all, they are supposed to involve technology that is far along in its development and from which industry will directly benefit. If industry isn't willing to make a substantial commitment,



## PARTY OF ONE

then that begs the question of whether the technology is truly ready for 'prime time'. By law, the private sector is required to provide 50% of the funds for demo projects.

But that was never the case with FutureGen. The Department of Energy argued from the start that it wasn't a true demonstration but a "test bed", more of a way to research new technologies. As a result, the politically potent coal industry had to contribute just 20% to the project — later raised to 26%. Last year, however, as the costs kept growing, the agency pressed the coal companies to pay 50% of the increases above the current \$1.8 billion price tag for the project. (The original cost estimate was \$950 million, but the two figures may not be comparable because they treat inflation differently.) The coal companies refused to budge, according to the agency, and they wanted to fund their share not with cash, but with debt leveraged against any money the plant would eventually make by selling electricity. This was apparently too galling even for the Department of Energy, and Secretary of Energy Samuel Bodman pulled the plug. Until that point, the agency had been negotiating in an attempt to find ways to continue the politically sensitive project it had touted so highly.

Not surprisingly, the main reaction in Congress has come from the congressional delegation from Illinois, the state where the FutureGen plant was to have been built. Indeed, the coal industry did everything possible to stoke the local politics. The site was chosen by industry, which announced the location in December, despite the Department of Energy's strenuous objections that doing so was premature. The Illinois delegation is numerous and powerful, and on a bipartisan basis it is fighting to keep

the project alive. It's not clear whether anyone else in Congress will care enough to stop it. If nothing else, the delegation could prevent a final decision from being implemented until the next administration takes office, which could be headed by a President Obama from Illinois.

Ironically, while hoping for a home state saviour, some members of the Illinois delegation, including Senate Majority Whip Richard Durbin (Democrat, Illinois) are accusing the Bush administration of cancelling the project because Mattoon — the Illinois town that was to get the plant — won out over two potential sites in President Bush's home state of Texas. Representative John Shimkus (Republican, Illinois) said he personally pleaded with the president to save FutureGen, but was told "Bodman says the costs are too high".

All this, of course, begs the question of whether the project ever made sense to start with. Demonstration projects are often driven more by politics than by science. Small-scale, targeted demos can be a genuinely useful way to test out new technologies that are far enough along in their development stage. But FutureGen was to build an entire coal plant using fairly well tested technology and then add to it a mix of technologies, some of which may have needed more research, and some of which needed smaller, focused demonstrations. What FutureGen was really designed to show was that the administration and the coal industry were working on climate change, with the implication that it was too soon to force them to do anything.

Which brings us to FutureGen's fatal flaw, which it shares with many other energy technology projects. Research and development is a necessary, but hardly sufficient, component of an energy and climate policy. Without government steps to develop a market, such as a way to place a price on carbon, there is no reason for industry to pursue advanced technologies. Indeed, a rationale for the revised FutureGen proposal is that industry has no reason now to invest in carbon capture and storage technology, which makes coal plants less efficient and electricity more costly. The day after a House hearing last month to question the cancellation of FutureGen, the president gave a speech about climate change in which he basically attacked every current proposal to do more than just jaw-bone industry and fund scientists. In such a context, any energy demonstration project ought to be suspect. It will take a lot more political pressure to turn the administration's demo projects from coal black marks into diamonds. ■

**David Goldston is a visiting lecturer at Harvard University's Center for the Environment. Reach him at [partyofonecolumn@gmail.com](mailto:partyofonecolumn@gmail.com).**





# THE 3-BILLION-DOLLAR QUESTION

Can a state do what a country cannot, and transform the way stem-cell research is funded?

**Erika Check Hayden** reports on the California Institute for Regenerative Medicine.

**E**van Snyder could hardly believe what he was hearing. About four years ago, Snyder, a neuroscientist and paediatrician, was sitting in the Los Angeles living room of Hollywood mogul Jerry Zucker, the producer and director behind movies such as *Airplane!* and *The Naked Gun*. Zucker's wife, Janet, and a mix of fellow scientists, Hollywood heavyweights, political staff and patient advocates were also there. And they were throwing around numbers that astounded Snyder, director of a stem-cell research programme at the Burnham Institute for Medical Research in La Jolla, California. Snyder had been invited to the meeting to discuss what he thought was an idea to raise US\$3 million for stem-cell research. That seemed like an enormous sum. Yet as the meeting progressed, it became clear that he had been mistaken.

"They said, 'You're missing the point — the \$3 million is to be able to get this on the ballot and get it passed for a bill that's gonna be \$3 billion,'" he says. "And I thought, I really am in la-la land now. Words like 3 billion dollars don't come out of the mouth of a scientist."

If \$3 billion seemed like a dream four years

ago, it is now a reality that is changing not only the way science is done in California, but is resonating across the US biomedical landscape. Since August 2001, when President George W. Bush banned federal funding for research on human embryonic stem-cell lines created after that date, six other states have passed initiatives to fund such work using state money. But none of them is even remotely close to the scale of the California Institute for Regenerative Medicine (CIRM), the research agency brought into being after California voters passed a ballot initiative in November 2004. All the other states together have pledged just over \$1 billion for stem-cell work (see *Nature* 451, 622–626; 2008).

Now, after giving about \$260 million through 123 research or training grants to 22 institutions (see 'Top 10 CIRM-funded institutions', overleaf), the CIRM is about to distribute an additional \$262 million to finance a network of major buildings. Next week, its oversight committee will make final decisions on recommendations made by a working group about where the money should go. The state's universities stand to receive major boosts to their already

powerful biomedical infrastructures, and have been anticipating this 'major facilities' initiative more eagerly than any other since the birth of the CIRM. But even as the agency is changing California's scientific outlook, it is also facing pressure to prove its worth to voters — and to show that it can deliver the medical and economic benefits it promised in order to convince taxpayers to fund it in the first place.

## Grand plans

On an ocean bluff next door to the Salk Institute for Biological Studies in La Jolla, an eroded strip of unused road winds through a 3-hectare grove of eucalyptus trees. Overhead, seagulls fly through a chilly afternoon fog towards the shore, where paragliders are launching themselves off the edge of a cliff. If everything goes to plan, in two years the eucalyptus trees will be gone. In their place will be a 12,700-square-metre building housing 280 scientists, who will be able to look out from their lab benches and watch the gliders drift by. This site is the future home of the San Diego Consortium for Regenerative Medicine, one of seven 'CIRM institutes' likely to be awarded

J. P. MOCZULSKI/REUTERS



NEWS.COM

funding next week — and one that would not have happened without prominent local philanthropists such as John Moores, Malin Burnham and Irwin Jacobs.

Scientific institutions have long relied on philanthropy to some extent, but the CIRM was born of their generosity. Individual donors contributed the bulk of the \$26 million in donations that backed the \$35-million campaign for Proposition 71, the ballot initiative that created the CIRM. Private donors again came to the agency's aid after the initial state bond offering to fund the agency was held up by lawsuits filed by groups opposed to stem-cell research. And in 2006, philanthropists lent the CIRM \$45 million, \$14 million of which bankrolled its first round of grants before governor Arnold Schwarzenegger (pictured, left, in research mode) came to the rescue with a \$150-million state loan. The litigation finally ended last year, freeing California to begin selling bonds.

For the imminent major facilities awards, the CIRM required institutions to come up with matching funds — which led San Diego scientists, such as Snyder and Larry Goldstein, stem-cell research director at the University of California, San Diego (UCSD), to Moores' door a few years ago. Moores, a software magnate, gave the scientists an imperative: he would give them the money to transform that eucalyptus grove next to the Salk only if they would build it together. Moores saw a joint CIRM building, housing scientists from San Diego's four premier research institutions, as something more than a monument to science; it would be a point of civic pride for San Diego.

In 2006, Moores put up \$6 million, and the San Diego Consortium for Regenerative Medicine was born, comprising UCSD, the Salk and Burnham institutes and the Scripps Research Institute. On 5 April, the CIRM's working group recommended that the consortium be awarded the agency's second-largest facility grant — \$43 million — to help build a \$115-million shared building. The consortium is an example of how California's philanthropists have become heavily involved in shaping the agency — and how scientists, in turn, have tapped into the state's private wealth.

In exchange, Californian universities have benefited enormously. The CIRM working group recommended that Stanford University in Palo Alto could receive the agency's largest facilities grant, of \$47.5 million, to build a \$200-million building it has had in the works since 2001. About 65 kilometres north, architects for the University of California, San Francisco (UCSF), designed an elegant building cascading down the face of Mount Sutro

— the hill that towers over the university's main campus — with a 'green' roof and an innovative system of split-level floors. For its bid, Stanford garnered at least \$40 million from alumnus and Business Wire founder Lorry Lokey; UCSF hauled in a \$16-million donation from Ray Dolby, founder of Dolby Laboratories, and his wife. Altogether, California institutions have committed and raised \$495 million in matching and leverage funds towards the facilities awards.

### Luring scientists

When the new buildings are complete, perhaps as early as 2010, they will be stocked with some of the dozens of stem-cell scientists who have been lured to California by the CIRM. Some are high-profile hires. In 2006, Peter Donovan left Johns Hopkins University in Baltimore, Maryland, for the University of California, Irvine, and Martin Pera moved to the University of Southern California, Los Angeles, from the Monash Institute of Medical Research in Victoria, Australia. Pera's former colleague, Alan Trounson, became president of the CIRM last year. But the CIRM has also been a boon for scientists whose names are not so well-known — people such as Su Guo, a 37-year-old UCSF biologist, and a postdoctoral fellow working with her, 35-year-old Julio Ramirez.

At 9 a.m. on a Tuesday morning, Guo and Ramirez study a petri dish of mouse embryonic stem cells in a lab at San Francisco's J. David Gladstone Institutes. The space is so jammed with fume hoods, refrigerators and micro-

scopes that there is almost no room to move around, and five other scientists, some of whom work for CIRM-funded investigators, are jostling for space around Ramirez. But he

doesn't mind the crowd — he has been studying animal stem cells for only four months, and the others are helping him learn the ropes. "I can ask them anything," Ramirez says.

Before this year, Ramirez studied plant stem cells, and Guo studied brain development in zebrafish. Yet they are now working with mouse and human embryonic stem cells, thanks to a \$564,000 grant from the CIRM. Guo's grant was one of 72 that the agency gave out last February as part of the \$45-million Scientific Excellence through Exploration and Development (SEED) awards, designed to support scientists who have never worked with stem cells. With her grant, Guo hired Ramirez.

First, Ramirez will look for chemicals that have positive effects on fish brain cells that produce a neurotransmitter called dopamine. Then, he will test these drugs on human neurons derived from embryonic stem cells to see

**"We have to be an instrument of change."**  
— Alan Trounson

## CALIFORNIA'S STEM-CELL AGENCY: THE FIRST FOUR YEARS

### November 2004

Proposition 71 is approved by California voters, creating the \$3-billion California Institute for Regenerative Medicine (CIRM).

### December 2004

Businessman Robert Klein is named chairman of the CIRM board.

### February 2005

A coalition of anti-abortion and taxpayers' rights groups file lawsuits challenging the legality of Proposition 71.

### March 2005

Zach Hall is named interim president of the CIRM; he becomes permanent president later in the year.

### April 2006

The CIRM awards its first grants: \$12 million in training grants.

### February 2007

The CIRM awards its first research grants: 72 SEED grants totalling \$45 million.

### March 2007

The CIRM awards 29 comprehensive research grants totalling \$77 million.

### April 2007

On 17 April, Zach Hall announces that he will resign from the CIRM effective 30 April.

### May 2007

Litigation against Proposition 71 ends, clearing the way for the agency to sell its first bonds; it does so in October, to the tune of \$250 million.

### June 2007

The CIRM awards 17 shared research laboratory grants totalling more than \$50 million.

### December 2007

The CIRM announces that board member John Reed is recusing himself from the agency's oversight committee pending a review into potential conflicts of interest. The CIRM awards \$54 million in 22 new faculty awards. Alan Trounson takes over as CIRM president.

### May 2008

Decisions are expected on how \$262 million will be parcelled out in major facilities grants.



whether they might treat symptoms of Parkinson's disease, which are caused by a dopamine deficiency. Guo has good funding support from the National Institutes of Health (NIH) — about \$800,000 a year for her zebrafish studies. But her CIRM money is letting her do something she has always wanted to do — test her work in human cells. “I always wanted my work to be applicable to patients,” she says.

As Guo and Ramirez examine the cells under a microscope, there is disappointment. The day before, Ramirez had injected the cells with a plasmid — a piece of DNA — that codes for a green fluorescent protein, so today the cells should be glowing green. But they are stubbornly pale. Guo quizzes him: is he growing the cells using the right ingredients? Did he use the correct plasmid?

Culturing stem cells is sometimes more art than science, as Ramirez is discovering, and even when you follow all the rules, sometimes things still don't work. Even once Ramirez becomes expert at manipulating the cells, further obstacles could torpedo the project. “What we've proposed to do is quite daunting,” Guo admits. Which raises the biggest question about the CIRM: will scientists be able to deliver the results it promised?

### Delivering results

This is an urgent concern for the leaders of the CIRM, because it won the hearts of California voters by saying it would produce cures for a number of debilitating diseases. Robert Klein, the chair of the CIRM board who masterminded the campaign behind Proposition 71, has a son with diabetes. The group formed to support the proposition, Yes on 71, paid for advertisements featuring patients such as the late Christopher Reeve, who implored voters to “stand up for those who can't” from his wheelchair. That ad also referred to research done by Hans Keirstead of the University of California, Irvine, who, in the run-up to the vote, gave talks around the state featuring a video of a formerly paralysed rat walking, thanks to transplants derived from stem cells. Keirstead's advocacy drew criticism from scientists who pointed out that he hadn't published his work in a peer-reviewed journal, and accused him of hyping the potential of stem-cell research. (He eventually published a paper: H. S. Keirstead *et al.* *J. Neurosci.* **25**, 4694–4705; 2005.)

Klein's coalition has now morphed into a group called the Americans for Cures Foundation, whose leadership includes former CIRM staff

members and a patient advocate, Don Reed, who has a paralysed son and has lobbied heavily on behalf of the CIRM. (Reed had previously lobbied the California legislature, successfully, for money that helped fund Keirstead's experiments.) The group's stated mission is “to support fellow advocates in the fight for stem-cell research and cures”.

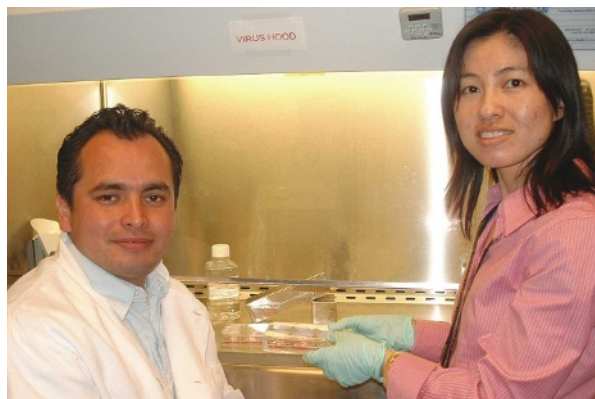
In 2006, the CIRM adopted a strategic plan pledging to meet high expectations with numerous and specific breakthroughs within ten years — including “clinical proof-of-principle that transplanted cells derived from pluripotent cells can be used to restore function for at least one disease”, as well as additional clinical trials of “therapies based on stem-cell research”.

Yet such timelines will be difficult, if not impossible, to meet. The Tufts Center for the Study of Drug Development in Boston, Massachusetts, for example, has estimated that the average drug takes 7 years to develop, at a cost of \$800 million, whereas biotechnology products require an average of 8 years and \$1.2 billion. The CIRM has only 10 years and \$3 billion. And, unlike drugs, stem-cell therapies have not yet been shown to work in humans, so they will almost certainly require a larger investment than a typical drug. No clinical trials of treatments derived from embryonic stem cells are yet under way, but Geron, a biopharmaceutical company in Menlo Park, California, has said it will ask the US Food and Drug Administration for permission to begin the first one this year.

Further lengthening the odds, the CIRM has already been criticized for exaggerating its role

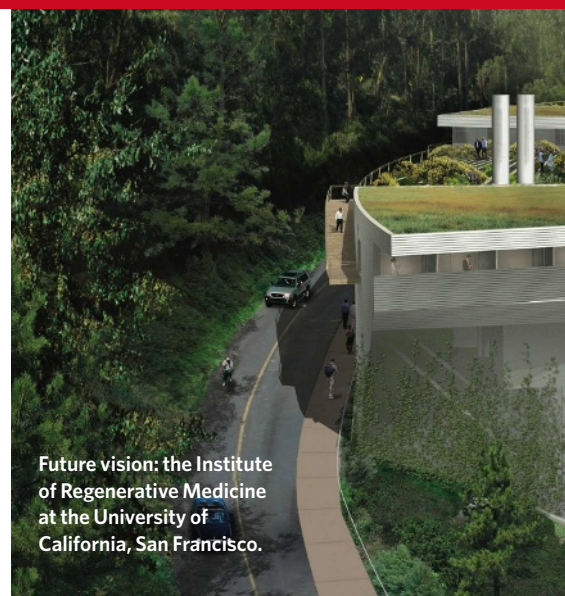
**“I always wanted my work to be applicable to patients.”**

—Su Guo



Julio Ramirez and Su Guo changed fields to work on human stem cells.

in research. Last month, for instance, the agency claimed credit for enabling a clinical trial to treat a group of blood cancers called myeloproliferative disorders. The trial is based on research described in an article published on 8 April (I. Geron *et al.* *Cancer Cell* **13**, 321–330; 2008).



Future vision: the Institute of Regenerative Medicine at the University of California, San Francisco.

On that date, the CIRM issued a statement claiming that an agency SEED grant to one of the authors — Catriona Jamieson at UCSD — had contributed to the article, and to the clinical trial. But the article was submitted last September — the same month Jamieson received her first SEED funding cheque. Still, the CIRM stood behind its statement, explaining that one of Jamieson's postdoctoral fellows had been drawing funding from an agency training grant since July 2006, and that the SEED grant was used to help the lab perform follow-up experiments requested by reviewers. And Jamieson calls the CIRM's funding “vital and instrumental”.

### Instrument of change

Agency leaders are also trying to shorten and cheapen the clinical-development process. For instance, they require CIRM institute applicants to describe how they plan to move discoveries into clinical trials. And next year, the CIRM plans to award up to \$20 million per team in ‘disease team research’ awards, which will

fund groups of researchers and doctors to try to move into clinical trials. “We have to be an instrument of change,” says Trownson. “Because if every single stem-cell treatment costs \$800 million, we're just never going to get there.”

The CIRM's structure is supposed to help keep it focused on cures. For instance, it gives patient advocates the power to make decisions about research right alongside scientists and businesses: on its 28-member board, 10 seats are set aside for patient advocates, 12 for Californian research institutions and universities, and 4 for businesses. But the CIRM's structure has, at times, seemed to hamper its own mission.

That was painfully evident at a meeting in January, when one doctor found himself begging for funding from 13 board members who were competing directly against him for money.

The doctor, Bertram Lubin, is head of the Children's Hospital Oakland Research Institute

UCSF

E. CHECK HAYDEN





(CHORI), a hospital-affiliated entity that draws about \$48 million in NIH funding a year. Lubin says that he applied last October for a \$5-million facilities grant to build labs where CHORI scientists could perfect and expand their method for using human-cord blood cells to treat sickle-cell anaemia and thalassaemia. The labs would also allow them to study whether placental cells could cure both diseases, which are blood disorders that disproportionately affect African Americans and other ethnic minorities. Because CHORI already had good clinical and preclinical data, and was asking for far less money than other institutions, Lubin's friends at other universities thought he was a shoo-in. So he was stunned when the CIRM's peer reviewers gave CHORI one of the lowest scores in the entire competition.

### Conflicts of interest?

The public review of Lubin's proposal stated that reviewers found several faults with it, including a lack of work on embryonic stem cells, no clear vision for integrating the pre-clinical and clinical work, and the fact that the "clinical program was reviewed as only incrementally advancing the field... In summary, reviewers felt these proposals would be more appropriate for individual translational or clinical research programs," the review stated.

To Lubin, it seemed like a classic example of how clinical-research proposals often fare poorly with peer-review panels composed largely of basic scientists. Yet the CIRM was supposed to be different. "This was supposed to be taxpayer dollars for cures," he says. So he decided to make his case directly to the CIRM board. But at its 16 January meeting, board members first argued over whether Lubin should be allowed to speak, and then debated whether he should be given three or ten minutes to make his comments. Finally, Lubin was allowed to speak in a three-minute comment. "[This] would bring cures to patients in this state within the first year or two, and when you report on what CIRM has done, this will be a major accomplishment," he argued.

Lubin seemed to win over the patient advocates on the board, such as Jeff Sheehy, a communications officer at UCSF, who represents people living with HIV/AIDS. "We're going to make a lot of rich people richer. Why don't we cure somebody?" Sheehy said at the time, according to a meeting transcript. But when the board voted on CHORI's application, it was defeated, ten to five; four of those who voted against it represented universities that had applied for facilities funding.

Klein argues that Lubin's application was trying to fit a square peg into a round hole. "He was trying to use a facilities grant process to fund a clinical grant," Klein says. But Lubin sees it differently. "We're not in the 'in' crowd," Lubin says. "So a project that was really going to go into patients was essentially triaged."

The episode is only one in a series of incidents that have raised questions about the wisdom of putting the institutions that benefit from the CIRM in charge of governing it. Last December, for instance, CIRM board member John Reed, president of the Burnham Institute, said he would recuse himself from the board pending an investigation by California's Fair Political Practices Commission. The commission is investigating a letter that Reed sent last August, excoriating CIRM staff for excluding

a Burnham adjunct professor from a grant competition. Ten grant applications had to be removed from the same competition because they included letters of support from members of the CIRM board. Both situations were inappropriate, CIRM staff said, because board members are not supposed to influence decisions that affect their own institutions.

That is not the way that Klein hoped the CIRM would be a model when he drafted Proposition 71. And he has far bigger dreams for the agency, ones that hinge on working the

kinks out in the California system first. Klein is looking to use the CIRM as a model to rethink the way Congress funds the nation's biomedical research through the NIH

**"We're going to make a lot of rich people richer. Why don't we cure somebody?"**

— Jeff Sheehy

— he would like to see stable funding for many years, rather than the current system of year-by-year appropriations.

Improbable as that hope may be, Klein cites baby steps that he thinks make it less unlikely. For instance, before the CIRM, no US government had sold bonds to finance research; now, not only has California done so, but Klein is lobbying the federal government to switch from taxable to tax-exempt bonds for this purpose. Klein says he has powerful friends on Capitol Hill, including Democrats Nancy Pelosi of California, speaker of the House of Representatives, and Harry Reid of Nevada, the Senate majority leader. In 2002, Pelosi and Reid supported Klein's quest to convince both houses of Congress to pass a unanimous consent measure, requiring the support of every member of the House and Senate, to fund \$1.5 billion in diabetes research in the waning days of that year's congressional session. Never before had a unanimous consent measure been used to fund medical research; Klein says the victory "demonstrates that patient advocates have the potential to do some remarkable things".

At 62, Klein seems unwilling to let go of the agency he brought into existence; there are some who speculate that he might be laying the groundwork for a political campaign in the future. And although it seems improbable that Klein could change the way Congress does business, four years ago, there were many — including people such as Evan Snyder, sitting in that Los Angeles living room — who felt the same way about Proposition 71. And next week, \$262 million will show just how wrong they were.

Erika Check Hayden is a senior reporter for *Nature*, based in San Francisco.

TOP 10 CIRM-FUNDED INSTITUTIONS		
1	Stanford University, Palo Alto	\$41,389,000
2	University of California, San Francisco	\$29,669,000
3	University of California, Los Angeles	\$23,350,000
4	University of California, San Diego	\$19,854,000
5	University of California, Irvine	\$19,615,000
6	Burnham Institute for Medical Research, La Jolla	\$16,538,000
7	University of Southern California, Los Angeles	\$15,205,000
8	Salk Institute for Biological Studies, La Jolla	\$14,244,000
9	University of California, Davis	\$13,526,000
10	J. David Gladstone Institutes	\$9,621,000

See Editorial, page 1.





# Quantum all the way

How does our classical world emerge from the counterintuitive principles of quantum theory? Can we even be sure that the world doesn't 'go quantum' when no one is watching? **Philip Ball** talks to the theorists and experimentalists trying to find out.

**K**eith Schwab builds bridges. By most people's standards, they are very small bridges indeed: around 8 thousandths of a millimetre long and 200 millionths of a millimetre wide, visible only under a microscope. But to Schwab's eye, these objects are huge. That's because he is hoping to see them behave according to the rules of quantum mechanics — rules that allow for bizarre, counterintuitive behaviour such as being in two places at once. Quantum mechanics is generally thought to govern objects such as individual atoms, not lumps of matter like these bridges, which contain tens of billions of atoms.

It is an ambitious goal. But Schwab, based at Cornell University in Ithaca, New York, is one of several experimentalists seeking to probe one of the great conundrums of modern physics: the quantum–classical transition. Here, the

fuzzy quantum world somehow gives way to the solid, definite certainties of the everyday 'classical' world as we go up the scale from atoms to apples. If these experiments manage to confirm current theories of this transition, they could turn long-standing preconceptions about quantum theory on their head.

Early quantum theorists treated the quantum–classical transition almost as a kind of sleight of hand, something that had to be imposed on quantum mechanics to recover the familiar world. Now, however, there are strong signs that the transition can be understood as something that emerges quite naturally and inevitably from quantum theory. If that's so, it implies that 'classicality' is at root simply another quantum phenomenon. "There's good reason to believe that we are just as much part of the quantum world as are the tiny atoms and electrons that sparked quantum theory in the

first place," says quantum theorist Maximilian Schlosshauer of the University of Melbourne in Australia.

Testing the new description of the quantum–classical transition involves experiments on systems ranging from photons to superconductors to microscopic vibrating beams. These efforts pose an extreme challenge to experimentalists, as they involve looking for very small effects on comparatively big things — rather like trying to detect the sag when a fly lands on San Francisco's Golden Gate Bridge. The effects very quickly get so small that many physicists believe it is absurd to try to see them. "One crowd says: 'Of course it will work — quantum mechanics says so,'" says Schwab. "The other says: 'There's no way it will work — these guys are nuts.'"

Yet there's now some urgency to the search: understanding the transition is becoming

D. PARKINS



crucial for the emerging field of quantum information technology, which includes quantum computing and ultra-secure data encryption and transmission. Such practical applications may depend on an ability to hold classical behaviour at bay, sustaining the quantumness at big scales.

In part this is a technical challenge, which the experiments on the quantum–classical transition might help to address. But it is also a challenge to our understanding of the basic theory. Some physicists think that how quantum behaviour should be interpreted remains as unclear today as it was when Niels Bohr, Werner Heisenberg, Albert Einstein and others were developing (and arguing over) the theory. “This is an emotional issue,” says quantum technologist Chris Monroe from the University of Maryland in College Park. “Some insist that there is no problem; others insist that there must be an infinity of universes, each with their own classical description of a definite quantum state.” (see refs 1 and 2.) As explorations of the quantum–classical transition become more sophisticated, it may become increasingly feasible to put such ideas to the test, answering questions that until now have remained almost metaphysical.

### Where does the weirdness go?

To understand what the quantum–classical transition really means, consider that our familiar, classical world is an ‘either/or’ kind of place. A compass needle, say, can’t point both north and south at the same time. The quantum world, by contrast, is ‘both/and’: a magnetic atom, say, has no trouble at all pointing both directions at once. The same is true for other properties such as energy, location or speed; generally speaking, they can take on a range of values simultaneously, so that all you can say is that this value has that probability. When that is the case, physicists say that a quantum object is in a ‘superposition’ of states.

Thus, one of the key questions in understanding the quantum–classical transition is what happens to the superpositions as you go up that atoms-to-apples scale? Exactly when and how does ‘both/and’ become ‘either/or’?

Physicists have proposed many answers to that question over the decades. But the most favoured one involves a phenomenon known as decoherence<sup>3</sup>, which was identified and elucidated in the 1970s and 1980s. Crudely speaking, decoherence is a sort of leaking away of quantum behaviour when a particle interacts with its surroundings — for example, when an atom or molecule collides with those around it, or when light bounces off it. All we are left with is a partial picture of the system

— a picture in which only a well-defined subset of macroscopic properties, such as position, are apparent.

Why those particular properties? In the 1980s, Wojciech Zurek, now at the Los Alamos National Laboratory in New Mexico, offered an answer. In effect, different quantum states have very different resistances to decoherence. So only the resistant states will survive when a system interacts with its environment. These robust states are those that feature in classical physics, such as position and its rate of change, which is associated with momentum. In a sense, these are the ‘fittest’ states — which is why Zurek and his colleagues call their idea quantum darwinism<sup>4,5</sup>.

Decoherence also predicts that the quantum–classical transition isn’t really a matter of size, but of time. The stronger a quantum object’s interactions are with its surroundings, the faster decoherence kicks in. So larger objects, which generally have more ways of interacting, decohere almost instantaneously, transforming their quantum character into classical behaviour just as quickly. For example, if a large molecule could be prepared in a superposition of two positions just 10 ångströms apart, it would decohere because of collisions with the surrounding air molecules in about  $10^{-17}$  seconds. Decoherence is unavoidable to some degree. Even in a perfect vacuum, particles will decohere through interactions with photons in the omnipresent cosmic microwave background.

In summary, decoherence offers a way “to understand classicality as emergent from within the quantum formalism”, says Schlosshauer. Indeed, this picture means that the classical world no longer sits in opposition to quantum mechanics, but is demanded by it.

It all sounds good in theory — but is decoherence real? Can we actually see quantum effects leaking away into the environment? Serge Haroche and his colleagues at the Ecole Normale Supérieure in Paris began to put the idea to the test in 1996 by studying bunches of photons held in a type of light trap called an optical cavity<sup>6</sup>. They passed a rubidium atom through the cavity in a superposition of two states. In one state, the atom interacted with the photons to cause a shift in their electromagnetic oscillations, whereas the other state caused no change. Then, when the experimenters sent a second atom through the trap, it was affected by the state of the photons induced by the first atom. But that effect became weaker as the photon field decohered, so the measurements made on the second atom depended on the

time delay since the passage of the first. In this way, Haroche and his colleagues could watch decoherence set in by altering the timings of the two atoms<sup>7</sup>.

### Sliding doors

The decoherence description shows that there is no abrupt boundary, no critical size, at which quantum behaviour switches to classical. And that blurry boundary itself shifts depending on how it is measured. “It is the choice of the measuring apparatus that defines whether a specific object is quantum or classical,” says Anton Zeilinger of the University of Vienna. His team provided an example of this nine years ago, when it demonstrated quantum interference between beams of  $C_{60}$  fullerene molecules<sup>8</sup> — hardly as classical as the footballs they resemble, but nonetheless big molecules that can be seen with an electron microscope. Interference — the addition or cancellation of overlapping waves — is in this case a purely quantum effect, and can’t be understood if the molecules are viewed as discrete particles. It

is possible only if the molecules are in a superposition of states — in several places at once. “If you scan with a scanning tunnelling microscope a surface to which fullerene molecules stick, you see the little soccer balls sitting there as classical objects,” says Zeilinger. “But if you choose our interference experiment set-up, they are

quantum mechanically delocalized.” In other words, he says: “The same object can behave as a quantum system in one situation, and as a classical system in another.”

This makes the fullerene experiment an ideal quantitative test of the theory of decoherence. In a subsequent experiment with even bigger  $C_{70}$  molecules, Zeilinger’s team saw that the molecules’ interference patterns gradually disappeared as the gas through which the molecules passed became more dense. In a denser gas there are more collisions, and thus more coupling to the environment, so the molecules decohere more quickly. The researchers found that the decay of quantum coherence and interference happened at precisely the rate predicted by the theory. They have now seen similar behaviour with even more massive molecules, including fluorinated  $C_{60}$  ( $C_{60}F_{48}$ ) and modified porphyrin molecules with a plate-like shape ( $C_{44}H_{30}N_4$ ) (ref. 9).

“The important point is that the conditions for an observation of interference effects can be precisely specified and quantified using theoretical decoherence models,” says Schlosshauer. “We no longer have to limit our-

“One crowd says ‘of course it will work’. The other says ‘these guys are nuts’.”  
— Keith Schwab

K. SCHLOSSHAUER



Maximilian Schlosshauer thinks that even 'classical' things are quantum in their own way.

selves to the assumption of a poorly defined yet fundamental divide between the quantum and classical realms."

Individual molecules still seem a long way from our classical world. But decoherence has also been watched in genuinely macroscopic objects: hoops of superconducting material. Superconductivity is an inherently quantum-mechanical behaviour, and electrical currents in a ring of superconducting material called a superconducting quantum interference device (SQUID) can be placed in a superposition of states that circulate in opposite directions<sup>10,11</sup>. These states can then be monitored by looking at the interference between them. In 2003, researchers at Delft University of Technology in the Netherlands showed that a related approach, using pulses of microwaves to excite oscillations in a superposition of superconducting current states, could also be used to investigate decoherence<sup>12</sup>. As in Haroche's experiments, decoherence shows up as a weakening of the correlations between oscillations of the superposition states as the time between pulses lengthens.

Superpositions in superconducting currents may still seem a little remote from the apparent absurdity of finding tangible objects, much bigger than atoms or molecules, in two places at once. But Schwab and his collaborators think that it should be possible to detect such quantum superpositions in 'rough' lumps of matter: their tiny resonating beams, which are examples of nanoelectromechanical systems (NEMSs). These objects are small enough that their vibrations should be governed by

quantum mechanics: they should be restricted to specific energy levels, and thus specific frequencies<sup>13</sup>. But the separation between energy levels is very small, so quantum behaviour will be blurred unless the resonators are kept very cold to eliminate thermal noise.

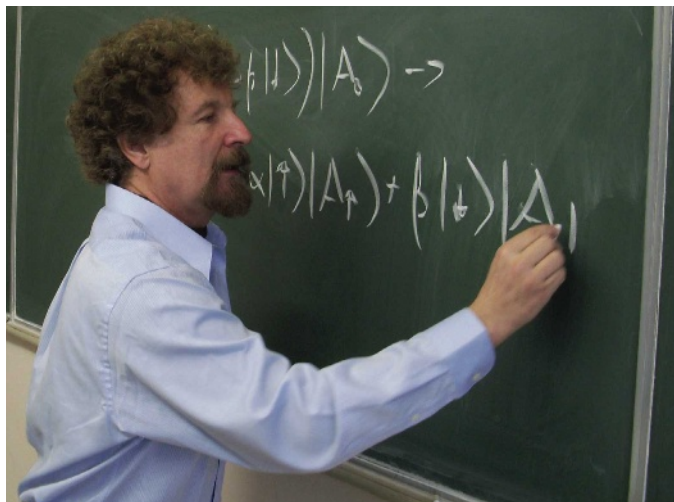
Schwab says that he thinks he is now very close to seeing a NEMS resonator in its lowest-energy (ground) quantum state. At temperatures of around 25 millikelvin, he and his colleagues have managed to restrict the vibrations to only the 25 lowest energy levels of the system, and he says he hopes to 'suck out' all the remaining excited states by active cooling, akin to the laser cooling used to draw heat from small clusters of ultracold atoms.

The Cornell researchers have proposed an experiment to search for superposition states in NEMS resonators<sup>14</sup>. As in other quantum systems, Schwab says, "as soon as you look for the superposition, you destroy it". So the superposition can only be inferred from its effect on another system to which the resonator is coupled. Schwab's

group will look at how interactions between the resonators and a nearby quantum bit or 'qubit' — a quantum device that can exist in two states, like the binary memory elements of a computer — induce decoherence in the qubit. This type of coupling will have a particular signature, says Schwab: coherence will reappear periodically at a rate equal to the vibration period of the beam. "That's the smoking gun," he says. This would reveal quantum effects in a system containing around ten billion atoms — much more than is feasible with molecular-interference experiments.

**"The same object can behave as a quantum system in one situation, and as a classical system in another."**  
— Anton Zeilinger

E. THOMMES



Wojciech Zurek proposes that decoherence selects for certain 'robust' states in an almost darwinian manner.

Even more ambitious, however, is a proposal by Dirk Bouwmeester of the University of California, Santa Barbara, and his colleagues to create superpositions in the position of macroscopic mirrors moved by the radiation pressure of a single photon in a superposition state<sup>15</sup>. Inevitably, this would involve extremely high-precision measurements: the researchers calculated that for a cube-shaped mirror measuring 10 micrometres in each dimension, with a mass of 5 trillionths of a kilogram and containing about  $10^{14}$  atoms, they would have to measure positions to within  $10^{-13}$  metres — close to the width of a proton, but nevertheless potentially feasible with interferometric methods.

### Smoke and mirrors

Bouwmeester hopes that it will soon be possible to test an alternative theory of the quantum-classical transition devised by Roger Penrose of the University of Oxford, UK. Penrose suggests that the 'collapse' of a superposition, rather than being a gradual affair resulting from environment-induced decoherence, is a rather abrupt event that is mediated by gravity. That is, it involves the emission of a graviton, the hypothetical fundamental quantum of the gravitational force, in much the same way that the decay of an excited molecule may happen via emission of a photon. He thinks that the cost in gravitational potential energy of keeping objects in a superposition becomes too great as objects get bigger, so that the objects 'go classical' on a definite timescale, which he estimates to be about a second or so for dust particles. Bouwmeester's mirror experiment should, if he can scale it up without losing sensitivity, be able to spot such a switch. "I am sceptical about this idea, but think it is worth testing," he says.



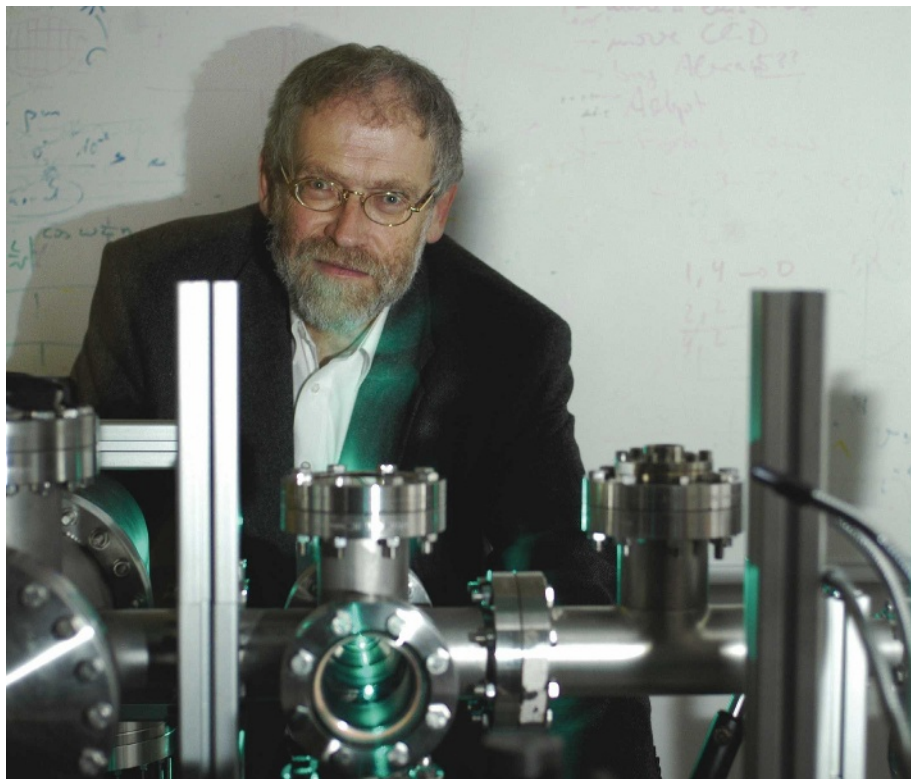
Understanding quantum states has become much more than an intellectual curiosity, because handling quantum data may hold the key to the future of information technology. By exploiting the extra degrees of freedom that superpositions offer, quantum computing could greatly increase computer power, at least for some computational problems. And encoding data in entangled quantum bits (the polarization states of photons, say) can provide a secure way to send information, as intercepting or reading the data would trigger irreversible quantum collapses that should be detectable. So understanding decoherence, the coupling of quantum systems to their environment and the fundamental nature of quantum measurement becomes crucial to manipulating information along quantum paths. "To build a quantum computer," says Schlosshauer, "the main challenge remains to shield the computer sufficiently from the environment to minimize decoherence while keeping it sufficiently open to allow its control from the outside. Research on quantum computing has inspired many ways of actively controlling decoherence and even of effectively 'undoing' its effects."

### Too much information

At first glance, quantum computing seemed to be sunk by the fact that superpositions can't be measured without destroying the 'information' they encode. But about 10 years ago researchers found that, as long as the background decoherence level in a quantum computer is small enough, quantum information stored in qubits can be retained by redundant encoding. "Several qubits are used to store just a single effective qubit of information," says Monroe, "and particular measurement schemes allow full recovery of quantum information in the face of decoherence."

One of the challenges in engineering these devices is identifying what is causing the decoherence. "The particular way to encode qubits will depend on the details of the decoherence source," Monroe explains. "Some decoherence sources are much easier to recover from than others."

If these efforts succeed in scaling up quantum computing from the mere handfuls of qubits demonstrated so far to truly macroscopic systems, the result will be not only a boon for computing itself but a forceful illustration of how quantum effects can survive at scales normally deemed 'classical'. "Some day we will have a roomful of equipment that will have to be considered a single quantum entity in the same way as a single atom is," says David DiVincenzo, who works on quantum



J. GODANY

In the quantum world, says Anton Zeilinger, what you see may depend on how the measurements are made.

computing at IBM's research labs in Yorktown Heights, New York.

On the other hand, maybe in pushing the envelope of the quantum-classical transition, quantum information technology will uncover some unanticipated, fundamental aspect of it.

"If in the process of building a large-scale quantum computer we find that decoherence always sets in at some level of complexity — a universal law that says quantum mechanics only goes so far — then that would be very interesting," says Monroe.

Zurek, meanwhile, sees a deeper significance in quantum information technology: he says it shows that information lies at the core of quantum theory. "Taking information seriously has profound consequences," he says. "Perhaps the most important is that there is no information without representation — a physical state of some object is needed to store information." Concepts such as decoherence and quantum darwinism make sense only because information matters, he says. "They all have to do with how information flows."

The decoherence description of the quantum-classical transition is not necessarily the end of the matter — it leaves unresolved some more fundamental questions about the

interpretation of quantum theory (see page 39). But at present, it seems a fair bet that what we think of as the classical world is really only the quantum world viewed through the lens of decohered states. "The conceptual leap would then be to conclude from this that quantum mechanics is truly universal," says Schlosshauer, "in the sense that everything, including us, is described by entangled quantum states." ■

Philip Ball is a consultant editor for *Nature*.

1. Buchanan, M. *Nature* **448**, 15–17 (2007).
2. Tegmark, M. *Nature* **448**, 23–24 (2007).
3. Schlosshauer, M. *Decoherence and the Quantum-to-Classical Transition* (Springer, Heidelberg/Berlin, 2007).
4. Ollivier, H., Poulin, D. & Zurek, W. H. *Phys. Rev. Lett.* **93**, 220401 (2004).
5. Blume-Kohout, R. & Zurek, W. H. *Phys. Rev. A* **73**, 062310 (2006).
6. Brune, M. et al. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
7. Raimond, J. M., Brune, M. & Haroche, S. *Phys. Rev. Lett.* **79**, 1964–1967 (1997).
8. Arndt, M. et al. *Nature* **401**, 680–682 (1999).
9. Hackermüller, L. et al. *Phys. Rev. Lett.* **91**, 090408 (2003).
10. Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. *Nature* **406**, 43–46 (2000).
11. Van der Wal, C. H. et al. *Science* **290**, 773–777 (2000).
12. Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. *Science* **299**, 1869–1871 (2003).
13. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. *Science* **304**, 74–77 (2004).
14. Armour, A. D., Blencowe, M. P. & Schwab, K. C. *Phys. Rev. Lett.* **88**, 148301 (2002).
15. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. *Phys. Rev. Lett.* **91**, 130401 (2003).

See Essay, page 39, and News and Views, page 50.

**"There is no information without representation."**  
— Wojciech Zurek

## Biopiracy rules hinder conservation efforts

**SIR** — The summits of the neotropical Guayana Highlands in Venezuela have a unique biodiversity that is under serious threat because of habitat loss resulting from climate warming. Although conservation studies are urgently needed, these are blocked by official bodies that will not grant permits for fieldwork in the region.

The bureaucratic process starts with the Venezuelan government's agency for science and technology, FONACIT, and involves a network of different organizations controlled by the ministry of the environment. These include the national tepui (table-mountain) commission, the biodiversity office, the office of indigenous affairs and the national institute of parks.

We have been involved since July 2005 in an international conservation project, funded by the Spanish BBVA Foundation, on the Guayana Highlands flora, in collaboration with several Venezuelan universities, research institutes and other organizations. It took us two years to obtain permits from FONACIT and for sample collection to be authorized. For unspecified reasons, the permits do not allow genetic studies, so molecular phylogenetic analysis is impossible. We are still trying to obtain approval from the office of indigenous affairs, but the people are reluctant to comply: they consider themselves owners of the summits, which are sacred lands to them.

The lengthy bureaucratic procedures have prevented scientific fieldwork in the Guayana Highlands for almost twenty years, when permissions to visit the summits for any purpose were suspended to avoid human disturbance and biopiracy (see below, 'Biopiracy: conservationists have to rebuild lost trust' *Nature* 453, 26; 2008). It is to be hoped that the situation may be reversed before it is too late to undertake suitable conservation strategies.

**Valentí Rull\*, Teresa Vegas-Vilarrúbia†**

\*Department of Animal Biology, Plant Biology and Ecology, Autonomous University of Barcelona, Faculty of Biosciences C1-339, Bellaterra, 08193 Barcelona, Spain

†Department of Ecology, University of Barcelona, Faculty of Biology, Av. Diagonal 645, 08028 Barcelona, Spain

## Biopiracy: conservationists have to rebuild lost trust

**SIR** — Working just south of Valentí Rull and Teresa Vegas-Vilarrúbia, we share the frustrations they describe on this page over obtaining permits — even though we only collect observations ('Biopiracy rules hinder

conservation work' *Nature* 453, 26; 2008). However, we disagree with the implications they draw for conservation.

Biopiracy is a matter of deep concern for indigenous peoples, who all too often have had their heritage abused. Practical conservation actions must adapt accordingly.

In 2001–02, a Japanese company controversially registered the common name of the fruit 'cupuaçu' (*Theobroma grandiflorum*) as a trademark. This aroused a great deal of alarm.

Brazil's national justice secretary, Romeu Tuma, is now asking Congress to consider a bill requiring foreigners — individuals, religious groups and environmentalists — to have a permit before visiting the Amazon (see <http://tinyurl.com/6kygfm>).

Contact with indigenous peoples is also contentious for other reasons. Whatever the merits of scientific exploration, indigenous peoples need to know how to evaluate researchers' requests and find out what is in it for them. After all, outside contact sometimes means massive cultural disruption — as with oil exploration in the western Amazon.

Conservation professionals can do much to improve the situation. They can start by acknowledging the rights of indigenous peoples, as recognized by Article 31 of the United Nations Declaration on the Rights of Indigenous Peoples (see <http://tinyurl.com/5okbd7>). Then, they can investigate indigenous people's needs through personal contact.

Practical actions that can be undertaken by researchers include providing Global Positioning System units and training in how to use them so that peoples can define their territories, helping to market traditional handicrafts or ecotourism, and building schools. Free, informed consent, obtained in advance through the proper channels, can ensure meaningful participation of indigenous peoples in project design and implementation. This should also minimize confrontations and the delays they cause.

Such actions are not a *quid pro quo*, but arise from a sense of justice. They can provide an important framework of trust for the ultimate conservation question: will indigenous groups manage their own lands, their own biodiversity, in sustainable ways?

Scientific needs are one small step on the way to that dialogue. We should not miss that step by focusing only on completing research projects.

**Mariana M. Vale\*, Maria Alice Alves\*, Stuart L. Pimm†**

\*Departamento de Ecologia, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, Maracanã, Rio de Janeiro 20550-011, Brazil

†Nicholas School of the Environment, Duke University, Box 90328, Durham, North Carolina 27708-0328, USA

## Spain should implement a model that's known to work

**SIR** — Your Editorial about the current state of Spanish science (*Nature* 451, 1029; 2008) suggests that 'A new Silver Age' is possible. Although Spain's research investment, at 1.1% of gross domestic product, is still below the European average (1.8%), the recent boost in the science budget has funded plans such as the Ramón y Cajal programme, intended to recover young scientists from abroad, and the creation of independent research institutes. But unless there are radical policy changes, the increased funding will not be sufficient to reduce the science chasm between Spain and other countries.

Blatant 'inbreeding' practices still thrive. Some 90% of Spanish scientists continue to occupy positions at the institutions where they obtained their PhDs, a situation that hinders the mobility and exchange of ideas that underpin strong science (see *Nature* 410, 14; 2001; *Nature* 411, 132; 2001). The Popular (conservative) Party, which governed from 1996 to 2004, brought in a national committee to replace the local committees that had previously been responsible for appointments. Although this maintained high standards for selection, it has failed to reduce academic endogamy. And although vacancies at most institutions are now publicized, this usually happens only after an in-house candidate has been unofficially chosen for the job.

A proposal by the current socialist government allows national evaluation of a candidate exclusively on the grounds of a detailed CV. However, that will still not solve the problem, as universities and research centres — which retain control of the final selection process — continue to favour their own candidates. Many of the Spanish regional governments have set up local evaluation agencies, often with different criteria: a positive report is mandatory when applying for a position in that particular region.

An even more flagrant practice favours candidates who are fluent in the local language, in addition to Castilian, irrespective of their scientific merits. This discriminates against non-local and foreign scientists trying to build a career in Spain. This is common practice and is in violation of the European Charter for Researchers (<http://tinyurl.com/42zbdg>).

Spanish policy-makers should stop reinventing the wheel in every four-year term. Instead, they should implement models that have worked for decades in countries with a long-standing tradition in science, such as the Max Planck institutes in Germany or the Medical Research Council in the United Kingdom.

Increased funding must be accompanied



by changes that, so far, Spanish politicians have ignored or failed to tackle.

The creation of a ministry of science and innovation and the appointment of a respected scientist, Cristina Garmendia, to head it, give us hope for a change in direction.

Unless the necessary changes are implemented, the careers of Spanish scientists, particularly the younger ones, cannot flourish.

**Rodrigo J. Carbajo\***, **José Luis Neira†**,  
**Rosa Farràs\***

\*Laboratorio de Biología Estructural,  
Centro de Investigación Príncipe Felipe,  
Avenida Autopista del Saler 16,  
46012 Valencia, Spain

†Instituto de Biología Molecular y Celular,  
Edificio Torregaitán,  
Universidad Miguel Hernández,  
Avenida del Ferrocarril s/n,  
03202 Elche, Alicante, Spain

## Spain: leading role of scientists is heartening

SIR — In your Editorial 'A new Silver Age?' (*Nature* 451, 1029; 2008), you highlight the urgent need for reforms in the organization of science, after science funding in Spain has been increased during the past four years. The socialist government has since been re-elected. What changes is this likely to bring?

Scientific and academic interests will be served by the newly created ministry of science and innovation, headed by biologist Cristina Garmendia, who runs a biotechnology company. The new government's composition is heartening news for scientists — for example, the minister of health is a professor of physiology, the minister of industry is an economist with a strong interest in science, and even the minister of the interior is a professor of chemistry.

Leading up to the general election of 9 March, the socialist party held out the carrots of more funding and structural reforms for science. Implementation of their programme will modernize a system that stems from one designed by the first socialist government in 1982, which first introduced grants and evaluation procedures.

Thirty years ago, I wrote an article about science in newly democratic Spain (*Nature* 274, 8–9; 1978). The situation has changed radically since that time, and it promises to continue to do so under a forward-looking stewardship.

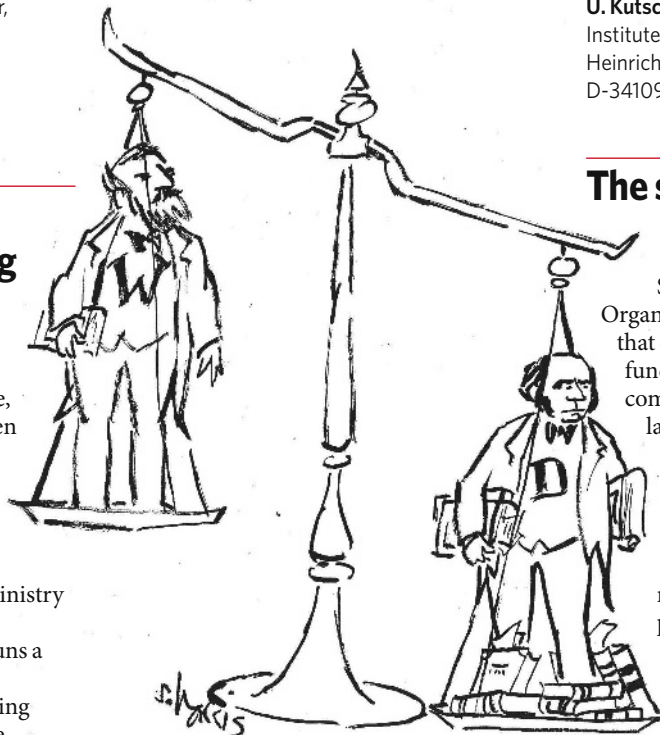
**Pere Puigdomènech**

Centre for Research in Agricultural Genomics,  
CSIC-IRTA-UAB, Jordi Girona 18,  
08034 Barcelona, Spain

## Darwin-Wallace principle of natural selection

SIR — In their Correspondence 'Celebrations for Darwin downplay Wallace's role' (*Nature* 451, 1050; 2008), G. W. Beccaloni and V. S. Smith question why Alfred Russel Wallace's achievements have been overshadowed by those of Charles Darwin, despite their discovery together of natural selection and its significance for the transformation of species (C. Darwin & A. R. Wallace *J. Proc. Linn. Soc. Lond.* 3, 45–62; 1858). I think the reasons for this are threefold.

First, Darwin's 1859 book *On the Origin of Species* describes the theory of descent with



modification by means of natural selection in much more detail than is found in his short essay with Wallace, published the previous year. The book became a bestseller and was translated into many languages. *Nature's* archives reveal the immediate impact of Darwin's monograph — see, for instance, T. H. Huxley's anniversary Editorial ('The coming of age of *The Origin of Species*' *Nature* 22, 1–4; 1880), but this made no mention of Wallace's contribution.

Second, Wallace had always acknowledged the priority of Darwin with respect to their joint discovery published in 1858. He used the term 'darwinism' as a synonym for 'the darwinian theory of natural selection' and popularized it (A. R. Wallace *Darwinism* Macmillan, London, 1889). To my knowledge, 'wallaceism' is a term that has never been coined.

Finally, Wallace was heavily involved with spiritualism by the 1860s. He confirmed his

belief in miracles and defended so-called supernatural phenomena, such as 'table-tapping', for the rest of his long life. This seriously undermined his credibility as a scientist, and cast a shadow over his brilliant theoretical work of 1858 on the struggle for existence in wild animal populations.

What can we do to rehabilitate Wallace and to acknowledge his important contributions to evolutionary biology? The 'Darwin-Wallace principle of natural selection' could be substituted for the old-fashioned 'darwinism', which smacks more of a political ideology than a modern scientific theory. This simple change in terminology might restore balance to the Darwin-dominated view of the history of the life sciences.

**U. Kutschera**

Institute of Biology, University of Kassel,  
Heinrich-Plett-Strasse 40,  
D-34109 Kassel, Germany

## The status of science in Muslim nations

SIR — The decision made by the Organization of the Islamic Conference, that only countries committed to donating funds will be allowed to join its standing committee on science and technology, is laudable and newsworthy. However, it hardly justifies the News in Brief headline 'Muslim nations raise status of science' (*Nature* 452, 517; 2008). A News Feature in the same issue reports the establishment of a research centre in Lisbon, made possible by the donation of half a billion euros (US\$782 million) by a single Portuguese philanthropist ('Navigating new waters' *Nature* 452, 528–529; 2008).

Nations are made up of individuals as well as governments. If governments in some Muslim states are not supporting science, why don't a few wealthy individuals step in? And if they did, would competent researchers, securely employed in a scientifically vigorous milieu, be willing to take up residence in the organization's member countries?

As your News Feature suggests, good scientists will go to Lisbon only if conditions are better than those elsewhere. The same applies even more forcefully to the Islamic countries, because conditions conducive to intellectual autonomy and scientific progress require more than a mere injection of capital. In Portugal, democracy needed to mature after the bloodless 'carnation revolution' in the 1970s ended a long spell of dictatorship.

**K. Razi Naqvi**

Department of Physics,  
Norwegian University of Science and Technology,  
N-7491 Trondheim, Norway

## COMMENTARY



# Making the grade

International testing that is used to predict the grim future of US science and technology is being vastly misinterpreted, say **Hal Salzman** and **Lindsay Lowell**.

It's a familiar story. Children around the world have been tested, and the United States is in trouble. A US Department of Education report<sup>1</sup> from March concludes that "without substantial and sustained changes to the educational system, the United States will relinquish its leadership in the twenty-first century". The panic plays out in countless newspaper articles and policy reports, recently leading to legislative responses such as the America COMPETES Act, which contains a list of measures to boost average mathematics and science test scores.

A country's place in the new global economy is, according to these reports, determined by its rank in the maths- and science-score hierarchy. Following this reasoning, one would conclude that the US economy is threatened not only by Japan and South Korea, but also by Finland, Singapore, New Zealand and the Czech Republic. The rankings that engender these fears are primarily based on two tests administered to middle- and high-school students since 1995: the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS).

Improving education should be a priority for the nation, but erroneous interpretations of international test scores may drive economic and competitiveness policy in the wrong direction. When we consider that education testing shows formidable US strength as the largest producer of top-scoring students alongside a significant problem at the bottom, the threat to future competitiveness seems to be something quite different from the headlines<sup>2</sup>. Caution is needed so we neither create policies that

overstock the science and technology workforce nor unthinkingly implement the education and social practices in other high-scoring countries. A full grasp of the meaning of test-score differences should lead the next president to address education and competitiveness problems more effectively than the recent America COMPETES legislation, which is now languishing for a lack of funding. Focusing the great consternation about education on real rather than imagined problems requires a careful assessment of the evidence.

## Lagging behind?

It is misleading to gauge the relative position of the United States in the world based on a simplistic ranking of its students' test scores. This is much like measuring shoe size to predict runners' future race times while ignoring their past performance. There are substantial methodological limitations in using these tests to compare nations, including reporting 'rankings' that are based on minute differences that are not statistically significant<sup>3</sup>. For example, when considering statistically significant differences, national test scores can be clustered into three meaningful levels and the United States consistently ranks in a middle group on maths and science while being top ranked in civics<sup>4</sup> — the study of citizenship and government. Overall, about one-fifth of other nations rank better and two-fifths rank underneath the United States.

Still, average performance tells us nothing about the distribution of students with the very best test scores. In maths and science, when looking at average scores, the United States is outranked by countries such as Finland and South Korea. But the rankings change when we examine the percentage of students who perform at the top, those most likely to be tomorrow's innovators. The South Korean average places it in the top-ranked group of nations, yet its relative proportion of top performing

students is 30% lower than that of the United States. In fact, the United States has a higher percentage of top-performing students than 5 of the 14 others in the top-ranked group of countries with high average scores.

Moreover, it would seem inappropriate to consider the United States, a country with a population of more than 300 million, in competition with Singapore, a country of 4.5 million, or with even smaller New Zealand. The economies in these countries range from a gross domestic product (GDP) of \$124 billion in New Zealand to \$236 billion in Finland, compared with the \$14-trillion GDP of the United States. Perhaps a more apt comparison would be Massachusetts with a population of 6.4 million and

a gross state product of \$338 billion, or Colorado with 4.8 million residents and a \$230-billion state product. Although the top group also includes economic powerhouses South Korea and Japan, which come in at under a fourteenth and less than a third, respectively,

of the size of the US economy, for the most part it makes more sense to compare US state economies with other countries because it is Massachusetts or California that is 'competing', for example, with Singapore in developing their biotech industries.

If, as we argue, average test scores are mostly irrelevant as a measure of economic potential, other indicators do matter. To produce leading-edge technology, one could argue that it is the numbers of high-performing students that is most important in the global economy.

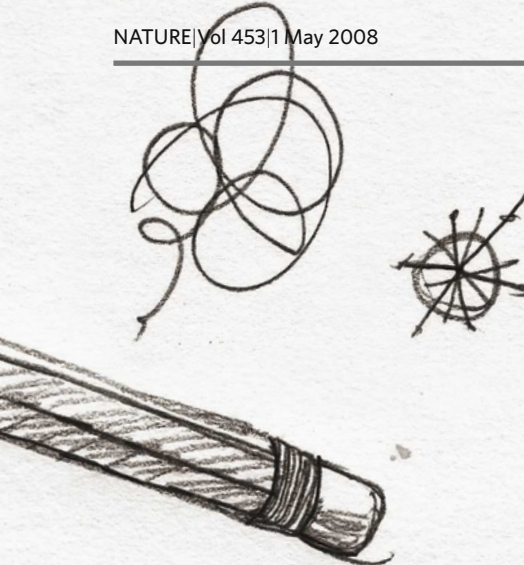
These are students who can enter the science and engineering workforce or are likely to innovate whatever their field of study. Remarkable, but little noted, is the fact that the United States produces the lion's share of the world's best students (see graph opposite).

At the same time, low-performing students can hamper productivity and here, unfortunately, the United States also stands out. The United States produces more than one million

**"Average test scores are largely irrelevant as a measure of economic potential."**

C. EMIABATA





low-performing maths and science students each year, more than any other country in the Organisation for Economic Co-operation and Development except for Mexico (see graph overleaf). Although programmes to improve education for low-performing students and schools are included in the various policy reports, they are shunted to the background when the headlines focus on increasing the numbers of those at the top, and overall seem to carry little weight when they are diluted as part of a long laundry list of recommendations.

### Market maths

Without a doubt, science, maths and technology education is needed in today's society, whether for its citizens to understand enough to participate in public debate or just to operate the technology of everyday life. However, some argue for more advanced courses as if they want to prepare all students to be scientists or engineers. We believe that there is something fundamentally wrong with such an approach.

History suggests that policies designed to stockpile scientists and engineers are counter-productive. The space race is typically cited as a success story of American technological prowess, but less often discussed is the impact of the workforce build-up on US engineering and science in the years that followed. Following a spike in the numbers of science and engineering college graduates in the late 1950s and early 1960s, a spectacular bust followed that led to high unemployment in these fields. For many years afterwards, fields such as physics were thought of as poor career choices<sup>5</sup>. Similar boom-and-bust cycles have continued for the past four decades, in engineering, in information technology (IT) and in science.

When demand is translated into

increased salaries and job openings, students respond. When the IT industry was growing, the number of graduates in computer science kept pace, doubling over six years. Following the collapse of the IT industry bubble, the number of graduates fell by 17% between 2003 and 2005. Employment in this field is just now reaching the levels of the boom years but, with little prospect of rapid growth, students seem to be wise in choosing other fields. Or, consider petroleum engineering. This is an industry that has had slow growth for two decades and, correspondingly, undergraduate enrolments declined 85% during that period, and master's programmes instead attracted students from areas of the world with fast-growing oil exploration. Today, 75% of US master's graduates in petroleum engineering are foreign students on temporary visas. Now, the US industry has a real need for more engineers because of increased demand for oil and new exploration coupled with 20 years of minimal hiring and an ageing workforce. The oil industry has responded by increasing entry-level salaries 30–60% over the past four to five years, far greater than in other fields. As a result, petroleum-engineering graduates have doubled in the past five years and freshmen enrolments, at Texas Tech University in Lubbock for example, have increased more than sixfold.

When supply far exceeds demand, the bust that follows reverberates for many years and discourages students even when demand does increase later. As Michael Teitelbaum of the Alfred P. Sloan Foundation in New York notes, not only is there no evidence of any widespread shortages but “substantially more scientists and engineers graduate from US universities than can find attractive career

openings in the US workforce”. Teitelbaum adds that this overproduction, which leads to increasingly longer postdocs in many science fields, makes our universities look more like a system to produce “a pool of low-cost research lab workers with limited career prospects than a high-quality training program for soon-to-be academic researchers”.

### Social choices

The beauty of brandishing a simple number or a few facts is that they fit in a single headline and focus the reader's attention. However, before we send teams of educators to discover the educational secrets of Finland, Singapore,

New Zealand, South Korea or Japan, we should do more study into the nature and context of their education systems. As the PISA report<sup>6</sup> notes, the tests do not evaluate schooling, per se, but the “cumulative impact of learning experiences... starting in

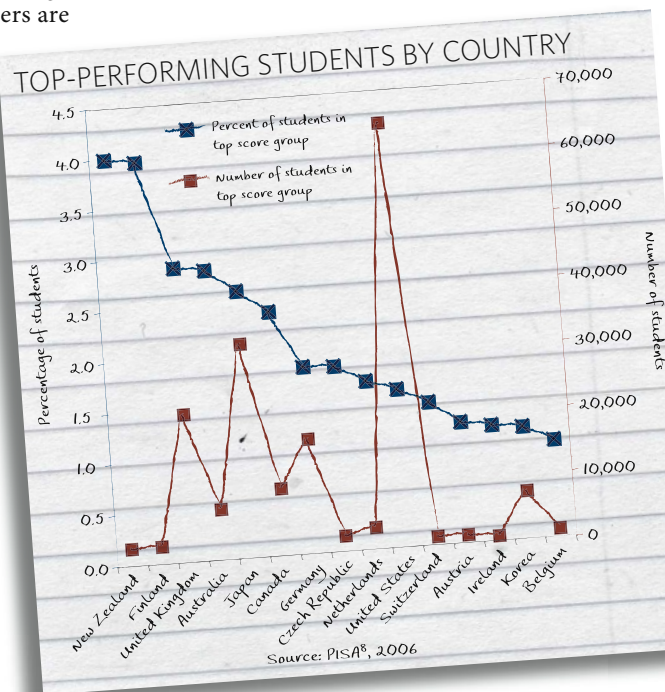
early childhood and up to the age of 15 and embracing experiences both in school and at home”. That is, much is made of a few select schooling practices in each country and recommendations are made to emulate them without considering what the effect is both on the lives of these children and on the economy.

Sending children to classes six days a week, extra preparation courses nights and weekends, and having a single examination that decides their fate, as is done in Japan, is not a choice most US parents would make. Nor is the social discipline in Singapore that seems to keep students on the straight and narrow path: death

for drug pushers, prohibitions on spitting and, for offences in between such as robbery after 7 p.m., at least 12 strokes of the cane. Although South Korea's spectacular economic rise is held in awe, and its tenfold-per-capita GDP increase over the past 20 years is widely praised, rarely noted is the close to 250% rise in the incidence of suicide over the same period, with suicide becoming a leading cause of death among young people. With South Korea are Finland and New Zealand at the higher end of the global rankings of test scores and suicide rates. No single factor is responsible for either high scores or suicides, but mental-health experts cite the pressure leading to one outcome as a factor leading to the other in many high-scoring nations<sup>7</sup>.

The future educational path for the United States should come from looking within the country rather than lionizing faraway test-score champions. Our analysis<sup>3</sup> of the

**“History suggests that policies designed to stockpile scientists and engineers are counter-productive.”**







data suggests two fundamental problems that require different approaches. First, pedagogies must address science literacy for the large numbers of low-performing students. Second, education policy for our highest-performing students needs to meet actual labour-market demand.

In the United States, a decade's worth of international test rankings based on slender measures of academic achievement in science and maths have been stretched far beyond their usefulness. Perhaps policy-makers feel it is better to motivate policy by pointing to high-scoring Czechs with fear, instead of noting our high-scoring Minnesotans as examples to emulate. But looking within the United States may be the best way to learn about effective education. As the PISA authors emphasize in their report, 90% of the variance in the scores is within countries rather than between countries. Therefore, most of what one can learn about high performance is due to the variation in factors within the nation's borders. It would seem far more effective to transfer best practices across city and state lines than over oceans.

### Chasing tails

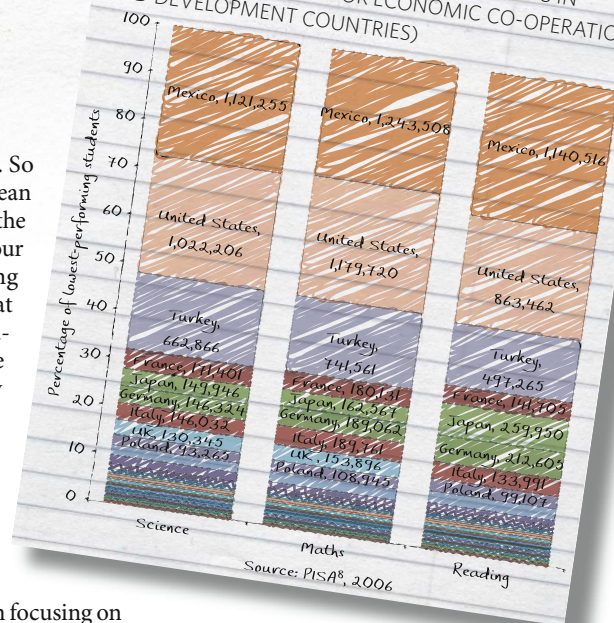
In America, little about the nation's condition can be gleaned from averages, whether by assessments of income or education. Our great opportunities as well as our great limitations seem to be accompanied by great disparities. It is these extremes, the tails at either end of the distribution, that require much more

discussion than the averages. So we need to look beyond the mean to consider size and quality of the workforce and the content of our education that will be driving innovation. In a country that has a long history of innovation and high productivity, we should start by looking at how our best schools educate top performers. It is unlikely that they do so by the types of education heralded in other countries.

Paying attention to the problems at the bottom is as important, if not more so, than focusing on the top. The most innovative technology has limited use if the more than 70 million workers without college degrees do not have the skills to use it effectively. The nation's low performers and schools should be a headline concern and the remedies are often to be found in schools only a neighbourhood or town away. It will be far more effective to take the best that America has to offer before seeking elusive and poorly understood practices found in a diverse collection of small countries around the globe.

As advocates of evidence-based policy, we argue that competitiveness and education policy should use the best available evidence as a guide and not be driven by impressions and rhetoric. Our analysis suggests that a better understanding of the education data will lead to better and, in many cases, different policy directions from those now being advocated. ■

LOWEST-PERFORMING STUDENTS BY COUNTRY: NUMBER AND SHARE OF TOTAL (15 YEAR OLDS IN ALL 30 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT COUNTRIES)



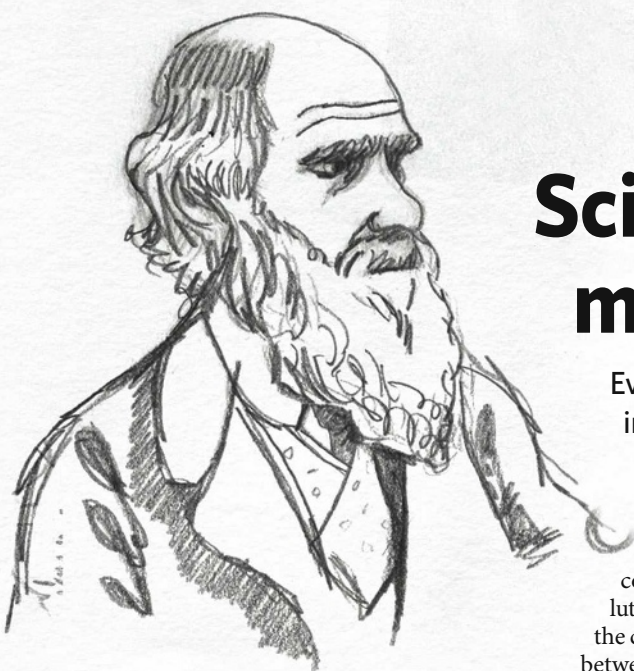
**"A better understanding of the education data will lead to better policy."**

Hal Salzman is at the Urban Institute, 2100 M Street NW, Washington, DC 20037, USA. Lindsay Lowell is at the Institute for the Study of International Migration at Georgetown University, 3300 Whitehaven Street, Washington, DC 20007, USA.

1. [www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf](http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf)
2. Lynn, L. & Salzman, H. *Issues in Science and Technology* Winter, 74-82 (2006).
3. Lowell, B. L. & Salzman, H. *In the Eye of the Storm: Assessing the Evidence on Science and Engineering Education, Quality, and Workforce Demand* (Urban Inst., Washington DC, 2007).
4. Boe, E. E. & Shin, S. *Phi Delta Kappan* **86**, 688-695 (2005).
5. Kaiser, D. *Social Res.* **73**, 1225-1252 (2006).
6. *Learning for Tomorrow's World: First Results from PISA 2003* (OECD, Paris, 2004).
7. Lim, M. *Asia's Ongoing Struggle with Suicide* International Affairs Journal at UC Davis (29 June 2007).
8. *PISA 2006 Science Competencies for Tomorrow's World* (OECD, Paris, 2008).

To discuss this article or any of our education material this week, visit <http://tinyurl.com/6ndqko>





# Science teaching must evolve

Evolutionary theory, study and knowledge moved on dramatically in the latter half of the twentieth century, but school teaching, curricula and teacher training are still in the primeval soup era, says **Andrew Moore**.

**W**e are approaching two important anniversaries for evolution: 2009 marks 150 years of Charles Darwin's and Alfred Russel Wallace's theses on evolution and 200 years since Darwin was born. While revelling in the memory of these world-changing events, we might also regret that many of the most fascinating and definitive examples supporting evolution — those made in the past four decades using gene-sequencing technology and bioinformatics — are largely absent from European secondary school curricula. Pupils are short-changed on molecular evolution, the best approach yet to investigating biology's central theory. The result: they leave school without fully understanding how well supported evolutionary theory is. Worse still, the understanding they have — based on the fossil record — is easy prey to specious arguments from anti-science movements.

The educational shortcomings are not for lack of information: countless excellent resources exist — albeit in a limited range of languages — and more are coming. In September of this year, molecular evolution will finally make it into all UK syllabuses for the A-level biology examination, a fitting early birthday present for Darwin. But that took some pushing, and more will be needed to extend the trend across Europe. Teachers need training to comprehend the material and its place in the pantheon of biological understanding. Moreover — and this goes for many topics in modern biology — the notion that 'molecular stuff' can be learnt later if desired must be overcome.

Molecular phylogenetics is routine science. In the early 1960s researchers were already

comparing gene sequences to infer evolutionary relatedness. Early work included the comparison of haemoglobin sequences between horses, pigs, cattle and rabbits, and between various primates<sup>1</sup>. In 1977, Woese and Fox overturned our understanding of the relationship between the fundamental kingdoms of life by comparing 16S ribosomal RNA sequences between two branches of organisms hitherto considered to be bacteria<sup>2</sup>. The result was a clear distinction between bacteria and archaea — still prokaryotes (organisms lacking nuclei and other intracellular structures) but much closer to eukaryotes (organisms with cells containing a nucleus) in important respects — and a revision of theories concerning the origin of the eukaryotic cell.

Today, more than 40 years after molecular-evolution research started, most schools across Europe still teach phylogeny based on comparative anatomy, embryology and physiology.

This works fairly well, but to preserve it as the paradigm to describe evolution in the face of a more robust approach is like continuing to

describe the universe and matter in terms of newtonian mechanics, neglecting relativity and quantum mechanics. In fact, it is worse, because newtonian mechanics is sound and consistent within certain frames of reference. Phylogeny based on similarity of form is fundamentally unsound because of the adaptation and convergent evolution witnessed in nature. Such purely descriptive methods represent a mould out of which schoolroom biology must break to become more contemporary.

## What is evolution?

Although the molecular mechanisms of evolution are mentioned in school curricula, molecular evolution itself is generally absent. In some countries it is only touched on in optional extras to the main syllabus. But why is this? Some involved in curriculum design argue that

there is already an overload of topics. But curricula are updated at intervals in all countries — presumably with new science. So, what is considered more suitable or important?

Speculations on the chemical origins of life are almost universally covered in school curricula under 'Evolution', despite the questionable relevance of the topic for evolution, and its rather uncertain scientific basis. At most it represents an opportunity to discuss the principles of disagreement and competing ideas in science. But it is far from sound evolutionary theory. Unfortunately, chemical evolution, or even the mere mutation of DNA or RNA bases, is often misinterpreted by teachers as molecular evolution.

Admittedly, curriculum design and update will always be problematic, but perhaps they would be less so if viewed from a new angle: the relevance of scientific advances to topics of societal interest. For here could lie a convenient justification for incorporating molecular evolution into secondary education. Mice are generally good models for human disease because their genes, and hence gene products, are closely related to ours. This knowledge results from genomewide sequence comparisons between mouse and man. Bioinformatics and genomics give us predictive power and the knowledge that the mouse is an invaluable model in biomedicine. In short, the use of animal models could be an excellent entry point for teachers, provided they have the right material.

## Language barrier

Europe certainly has accessible information on biomedicine, particularly genomics. The EuroGentest website, for example, hosts many links to educational projects and resources. However, they are unevenly distributed when it comes to language. Furthermore, there is a big difference between mere information

**"Pupils leave school without fully understanding how well supported evolutionary theory is."**

C. EMILABATA



and a teaching unit: teachers generally have no time or guidance to elaborate new course units from raw materials. Some initiatives from universities and research institutes (see 'Genetics education materials') offer in-service teacher training and exercise 'kits' on bioinformatics — another area generally neglected at school, but key to understanding genomics and molecular phylogeny. However, the coverage is small.

The language distribution of molecular evolution resources is even worse, although the Internet abounds with these too. Almost all are provided by institutions and organisations in the United States (see 'US educational resources'). Examples even include resources for pupils as young as 11–14 years. Searching for 'molecular evolution' on the website of the US National Science Teachers Association (NSTA) returns links to innumerable teaching resources, many of them free. In Europe, such resources are almost non-existent. Even in one of Europe's major languages — German — one finds little on the Internet besides a couple of in-service teacher training sessions restricted to their respective *Länder* (German states). Europe has no equivalent of the NSTA, but the Association for Science Education (ASE) comes closest. Browsing the publications list on the ASE website returns 160 entries. Only one is about evolution, and that is for primary schools. There is no mention of molecular evolution.

A developer of first-class teaching resources in molecular biology, the National Centre for Biotechnology Education (NCBE) in Reading, UK, will soon make a significant contribution for molecular evolution. Working with research scientists from several institutes, and with funding from the United Kingdom's Wellcome Trust, the NCBE is preparing "A birthday present from Mr Darwin": a teaching resource on modern concepts in evolution, to be circulated to UK schools and hosted on a freely accessible webpage. One of the 12 exercises, for example, involves studying the melanocortin-1 receptor (McR1) gene sequences from paleontological finds to deduce the probable coat colour of woolly mammoths.

The NCBE also coordinates the EU-funded project 'Volvox', the first concerted effort to translate and disseminate biology education resources in Europe. The few resources on molecular evolution must receive priority treatment for translation.

### Teaching the teachers

It is, however, worth asking how teachers will deal with the material. The theoretical basis of molecular phylogenetics should be simple to grasp. A useful, albeit imperfect, analogy is the tracing of modern European languages back to archaic language groups by studying 'mutations' in spelling and pronunciation. But preparation should be paramount. The Portuguese secondary curriculum, for example, states that evolution should be taught using all of the tools available. But equipping teachers with modern tools is rather left to chance, and many have trouble with ideas such as the molecular clock — a basic concept in molecular evolution. Even in the United Kingdom, recent findings suggest that many teachers are not scientifically capable of teaching evolution using modern approaches — a matter of teacher training<sup>3,4</sup>.

The problem in most countries starts here;

**"Many teachers are not scientifically capable of teaching evolution using modern approaches."**

### Genetics education materials

#### Some European organizations providing educational materials or training in molecular genetics and bioinformatics.

- Xlab, Göttingen: [www.xlab-goettingen.de](http://www.xlab-goettingen.de)
- Gläsernes Labor, Berlin: [www.glaesernes-labor.de](http://www.glaesernes-labor.de)
- École de l'ADN, Nîmes: [www.ecole-adn.fr](http://www.ecole-adn.fr)
- European Learning Lab for the Life Sciences at the European Molecular Biology Laboratory (EMBL): [www-db.embl.de/jss/EmblGroupsOrg/g\\_213.html](http://www-db.embl.de/jss/EmblGroupsOrg/g_213.html)
- Life Learning Centre, Turin: [www.lfctorino.it/index.php](http://www.lfctorino.it/index.php)
- DNA to Darwin project: [www.dnadarwin.org](http://www.dnadarwin.org) (coming soon)
- Volvox: [www.eurovolvox.org](http://www.eurovolvox.org)
- EMBO: [www.embo.org/scisoc/tw03biocomputing.pdf](http://www.embo.org/scisoc/tw03biocomputing.pdf)

### US educational resources

#### Some US websites providing educational resources on molecular evolution.

- University of Indiana molecular phylogeny education resources: [www.indiana.edu/~ensiweb/lessons/mol.bio.html](http://www.indiana.edu/~ensiweb/lessons/mol.bio.html)
- Molecular Evolutionary Genetics Analysis: [www.megasoftware.net](http://www.megasoftware.net)
- Public Broadcasting Service educators' pages: [www.pbs.org/wgbh/evolution/educators/course/session3/elaborate\\_a.html](http://www.pbs.org/wgbh/evolution/educators/course/session3/elaborate_a.html)
- Middle School Portal: <http://msteacher.org/epubs/science/science15/background.aspx>
- National Science Teachers Association: [www.nsta.org](http://www.nsta.org)

future educators are not properly exposed to molecular approaches and the shift in awareness that accompanies them. The problem continues in the form of inadequate in-service training. Textbooks will always be out of date and curriculum content arbitrary. But teachers can and should have formal annual updates in important areas of modern biology and biotechnology. At the Weizmann Institute of Science, in Rehovot, Israel, an experimental project supports teacher training with adapted primary literature<sup>5</sup>. With the help of scientists in various fields, two key research papers from top journals are selected annually and rewritten in accessible language. Certainly, there is no lack of universities in Europe with staff who support teachers part-time, voluntarily and unpaid; rather there is a lack of political recognition that this service must be paid for and expanded.

The absence of molecular evolution from secondary school biology teaching cannot be excused by claiming that it is 'molecular stuff' that can be learnt later. There is something more serious at stake: the erosion of public trust in Darwin's original theory of evolution by natural selection in the face of 'alternative theories' from the Intelligent Design movement. An article in *Science* recently demonstrated a moderate correlation between knowledge of genetics and acceptance of the theory of evolution among members of the general public<sup>6</sup>. No less than before, evolutionary theory needs to be buttressed by all the good science it can get, and there is no better place to start than in school.

Andrew Moore is manager of the Science & Society Programme at the European Molecular Biology Organization (EMBO), Meyerhofstraße 1, D-69117 Heidelberg, Germany.

1. *Evolving Genes and Proteins* (eds Bryson, V. & Vogel, H. J.) (Academic Press, New York, 1965).
2. Woese, C. R. & Fox, G. E. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
3. [http://www.iob.org/userfiles/File/JBE\\_archive/JBE\\_42\\_1\\_Cleaves.pdf](http://www.iob.org/userfiles/File/JBE_archive/JBE_42_1_Cleaves.pdf)
4. Williams, J. D. *Evolution Education Outreach* **1**, 87–95 (2008).
5. Falka, H., Brill, G. & Yarden, A. Teaching a Biotechnology Curriculum Based on Adapted Primary Literature. *Int. J. Sci. Edu.* doi:10.1080/09500690701579553 (2007).
6. Miller, J. D., Scott, E. C. & Okamoto, S. *Science* **313**, 765–766 (2006).



## BOOKS &amp; ARTS

# Mix and mash-up

An experimental musician explores how technology has transformed our cut-and-paste culture.

## Sound Unbound: Sampling Digital Music and Culture

edited by Paul D. Miller aka DJ Spooky that Subliminal Kid

MIT Press: 2008. 416 pp (plus CD).

\$29.95, £17.95

### Marc Weidenbaum

Not every disc jockey has their first book-jacket blurb penned by Stanford University law professor Lawrence Lessig, the public-intellectual face of copyright reform in the era of digital technology. Nor does every DJ have their second book introduced by Steve Reich, a leading minimalist classical composer. Nor is it common for a DJ to have their first two books published by the Massachusetts Institute of Technology (MIT) Press.

Then again, 'DJ' doesn't mean what it used to. No longer just a disembodied radio announcer (think of the iconic and elusive Wolfman Jack in the George Lucas film *American Graffiti*), the DJ has become a centre-stage cultural figure — performer, composer, remixer, sound artist and activist. DJ Spooky, born Paul D. Miller in Washington DC in 1970, is all of those and more. The experimental musician and producer has applied his mixing techniques to film (reworking D. W. Griffith's *Birth of a Nation*), to soundtracks (accompanying a 2001 issue of *Nest* magazine and both of his MIT books), and to the archives of esteemed record labels (the avant-garde label Sub Rosa for his MIT projects, and reggae label Trojan). He has exhibited at the Whitney and Venice contemporary art Biennials, and interacted with artists from Yoko Ono to death-metal drummer Dave Lombardo. A professional audio provocateur, Spooky revels in the twenty-first-century enthusiasm for artistic border-crossing, a phenomenon fuelled by rapid advances in technology that have transformed art and communication.

Investigating the nature of that transformation is the purpose of Spooky's latest book. *Sound Unbound* is a collection of essays that explores the ways in which our culture is built on sampling. His previous book, *Rhythm Science* (MIT, 2004), defined sampling as "creating with found objects". That description makes it easy to connect the dots from the ready-made art installations of Marcel Duchamp to the literary cut-ups of novelist William S. Burroughs, to the tape manipulations of The Beatles to the loop-based music-making that defines hip-hop. Today's media landscape is characterized



JASON LAVERIS/WIREIMAGE.COM

Disc jockeys such as DJ Spooky epitomize today's trend for sampling and artistic border-crossing.

to an unprecedented degree by cultural appropriation, genre bending, networked collaboration and high-tech craftsmanship.

Edited by Spooky, *Sound Unbound's* three-dozen essays, interviews and poems by various science, art and cultural commentators explore diverse if interrelated subjects. Bruce Sterling reflects on early technologies that failed to gain the public's support, precedents to the rivalry between VHS and Betamax video formats. In his tour de force *The Ecstasy of Influence*, originally published in *Harper's* last year, Jonathan Lethem treads the fuzzy line between plagiarism and originality and then exposes how his essay was built, like a model from a toy construction set, out of other people's words and thoughts. Composers Pierre Boulez and Steve Reich (along with Reich's wife and collaborator, the video artist Beryl Korot) submit to detailed interviews about their creative maturations. Musician Brian Eno contributes a compact, erudite history of bells (taken from the liner notes from his 2003 album, *January 07003*). Frances Dyson and Douglas Kahn struggle to locate a metaphor for navigating today's digital-media saturation. Erik Davis divines the spiritual roots and metaphysical future of dub music. And Naeem Mohaiemen and Ron Eglash focus on racial implications in rap music and computer circuits.

Spooky emphasizes how the present bleeds into the future, yet the book is deeply rooted in

the past. Jeff E. Winner provides a useful history of Raymond Scott, the composer and electronic innovator. Joseph Lanza shows how 'easy listening' prefigured 'ambient' as a celebration of background music. Ibrahim Quraishi likens ancient Islamic ritual music to modern-day music loops. And in an interview that might be mistaken as extraneous, long-time album-cover artist Alex Steinweiss talks about the packaging of music. Given that compact-disc sales are diminishing more quickly than the sales of music downloads are rising, his reminiscences take on the aura of a requiem.

Clearly applying his DJ skills to editing, Spooky layers seemingly incongruous material and lets the sympathetic overtones register with the reader. As a bricoleur, he is a little light on the mortar that binds the book: aside from a brief introductory essay, few connections are drawn between the chapters. Some bricks are less sturdy than others. An interview with electronic musician Moby has nothing to distinguish it from his countless other interviews. Pointillist, anecdotal texts by musicians Chuck D, Saul Williams and Daniel Bernard Roumain do not reflect their accomplishments in recordings and in concert. And Jaron Lanier, in the book's cranky closing rant, overemphasizes hip-hop's relationship with digital technology and under-acknowledges the impact of MP3 file-sharing on music sales (he suggests that the

record industry's decline is caused by today's music being "crummy").

The absence of biographies for the contributors is strange. It would have helped to learn that Daphne Keller — contributor of perhaps the most clearly articulated essay, describing the US legal system's adjustment (or lack of it) to the digitization of culture — is a product lawyer at Google. It would have been useful to know that Ken Jordan, who writes about digitally induced synaesthesia and co-authors an overview of networked collaborative art with Spooky, is both a founder of the entertainment website [www.sonicnet.com](http://www.sonicnet.com) and an editor of *Multimedia: From Wagner to Virtual Reality* (W. W. Norton, 2001). Readers might have liked to learn that many contributors are musicians — not just Eno, Roumain and Williams, but also Vijay Iyer, Pauline Oliveros and Robin Rimbaud (known as Scanner), among others.

The book's index is especially haphazard.

For instance, multiple references to rap group De La Soul and to drum-and-bass act 4 Hero go unindexed, whereas single-instance nods to soundtrack composer Jack Nitzsche and rock band Rush make the cut. The Boulez interview cites Max Mathews, the computer-science legend and namesake of the popular Max/MSP music software program, but he is not in the index because a translator mistakenly rendered his surname as "Mathieu".

The 45-track CD that accompanies *Sound Unbound* illustrates and parallels the book's central argument, locating a historical foundation for today's innovations. Explanatory material beyond the song listings would have been appreciated, however. Segues make unforeseen associations, such as when an Erik Satie orchestration blends into a Steve Reich woodwind piece, suggesting a distant precursor to contemporary minimalist composition. The set is heavy on excerpts from avant-garde stalwarts, serving as a primer both on early

pioneer composers (Edgard Varèse, Pierre Schaeffer) and on modern figures from the laptop era (Ryoji Ikeda, Carsten Nicolai). Evident in much of the music is how producers such as Spooky and Bill Laswell filter existing material through their own record collections and musical equipment; on several tracks, DJs take archival spoken-word recordings of Marcel Duchamp, Jean Cocteau, Antonin Artaud and others and set them against a groovy back beat.

DJs may have come a long way from the halcyon era of night-hawk broadcasters to the brave new world of mash-up-happy culture vultures, but one thing has remained — successful DJs need to know their audience. Perhaps the audience for MIT's publications expects just a little more rigour, a little more structure and a little more editorial rhythm. ■

Marc Weidenbaum is an editor and writer living in San Francisco, California. He writes about ambient and electronic music and sound art at [www.disquiet.com](http://www.disquiet.com).

## Exemplary epidemiology

### Modeling Infectious Diseases in Humans and Animals

by Matt J. Keeling and Pejman Rohani

Princeton University Press: 2007. 366 pp.  
\$65, £38.95

### Mark Woolhouse

Infections produce further infections. The implications of this simple observation have long intrigued theoreticians and confounded empiricists. It implies nonlinear dynamics, to use the mathematical jargon, and this makes it difficult to be intuitive about what will happen next, especially if the intention is to intervene. Expert opinion is often not up to the task; we also need the insights provided by mathematical models. These are being widely used to help understand the epidemiology of infectious diseases and to design control programmes. Models can support, add to and sometimes even overturn prevailing wisdom — think of malaria, AIDS, measles or foot-and-mouth disease.

In 1991, Roy Anderson and Robert May published the hugely influential *Infectious Diseases of Humans* (Oxford University Press). The subject has since advanced significantly, and *Modeling Infectious Diseases in Humans and Animals* meets the need for a new synthesis. Authors Matt Keeling and Pejman Rohani are mathematicians by training who have made important and original contributions to epidemiology, so they are well qualified to deliver an authoritative, comprehensive and up-to-date review.

Their book contains a guide to different models and provides worked examples of the insights that models offer, and of specific applications to real-world problems. They cover an impressive range of mathematical approaches, from two-line coupled differential equations through

event-based stochastic models to spatially explicit microsimulations, and many others. Their examples cover an equally wide range of infectious diseases, from measles in school children to sexually transmitted infections in koalas. In every case, there is a thoughtful description of the rationale for the model, the assumptions behind it, the types of question it can be used to address, how to implement it (helpfully supported by a website providing access to computer code), and what the model tells us.

With all of this to hand, is the reader fully equipped to become a modeller of infectious disease? Not quite. Modelling is more than a technical exercise. It also requires that the practitioner makes critical judgements at different stages of the process, notably design, parameterization, validation and prediction.

Model design is the first and most important step. Success depends on how well we pose the questions we want to answer, and how effectively we identify the essential biology and translate it into mathematical equations or computer code. Keeling and Rohani manage this effortlessly, but it is a difficult art to instil in others except by example. There are plenty of examples in their book that repay close attention: particularly the sections on seasonality and contact tracing.

The second step, and an active area in the field, is model parameterization. It is not a major theme of *Modeling Infectious Diseases*. It was once acceptable to run a projection through some data points and declare the model good enough. This is no longer the case. More powerful computers and software have increased the availability of sophisticated estimation techniques, often using bayesian methodologies.

The third step is validation — the extent to which we should believe, and sometimes act on,

the output of a model. Keeling and Rohani take a mathematician's view of this. Their book is punctuated by concise summaries of the insights drawn from the models, presented as robust conclusions. These are helpful in communicating key results but empiricists will often, rightly, demand something more. Ideally, this should include testing model predictions against independent data.

Prediction is a difficult task that we routinely undertake, for example, when making a decision about implementing disease-control measures. Such decisions must always involve some kind of model, even if it is only a mental one. Mathematical models have two huge advantages. First, they are transparent — the inputs, assumptions and logic are available for inspection, criticism and change in a way that is rarely the case for expert opinion. Second, models can be used to explore, *in silico*, the expected impacts of many more different control options than could ever be trialled in practice. Often, models will be the best evidence we have for our decisions.

Keeling and Rohani advocate, as strongly as I do, the use of mathematical models to help design disease-control programmes and they devote the final chapter to this topic. They recognize that modelling is a partnership between modellers and empiricists, including experts in the disease system of interest, providers of epidemiological data and those responsible for disease control. For that reason, I hope that the readership of *Modeling Infectious Diseases* will extend beyond existing and new devotees of this challenging and exciting discipline. Most medics, vets and health workers will never write a mathematical model themselves, but it is increasingly important that they are familiar with the work of those that do. ■

Mark Woolhouse is professor of infectious disease epidemiology at the Centre for Infectious Diseases, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JF, UK.



## EXHIBITION

# Measure for measure

**John D. Barrow**

Mathematics is more than numbers. It's the collection of all possible patterns, from crazy-paving to crystals, carvings and castles. The exhibition *Beyond Measure* at Kettle's Yard in Cambridge, UK, brings together patterns with a geometrical flavour in different media.

Many patterns we find attractive, others not. Indeed, the Artist Paul Klee remarked that emphasizing only the beautiful is like using a mathematical system that handles only positive numbers. *Beyond Measure* is an illuminating collection of exhibits that mixes both; the displayed objects combine aesthetic quality with functionality because they were created to illustrate, to teach, to decorate or to demonstrate.

Some exhibits hail from the past, such as wooden models of platonic solids, archimedean polyhedra and geological crystals. Some are wonderfully contemporary, like the complex building models (pictured) by London-based architects Foster and Partners, sculpted using a three-dimensional colour 'printer' that gradually builds up the solid structures by laying down hundreds of strata of thin, coloured laser-cut paper.

The diverse collection captures the nineteenth-century extension of geometry into geometries. The discovery of non-euclidean geometries dismissed the old belief that euclidean space was the only logical possibility, forcing mathematicians to accept a point that had been appreciated and demonstrated by artists for more than 400 years. This relativism unleashed a rush of innovative designs and new explorations of the geometry of space, from cubism to its mind-bending counterparts based on the geometries of the sphere, the cylinder and the cone.

Kettle's Yard curator Barry Phipps searches out geometry in unexpected places. Beautiful drawings by medical consultants show the steps in a surgical operation, annotated with reassuringly illegible medical script and scalpel guidelines drawn on the patient's arm. Collections of subtly coloured, crocheted representations of the complex surfaces of Calabi-Yau spaces are presented for string theorists to knit their brows over, alongside ball-and-stick models from the laboratories of eminent biochemists who imagined the structures of proteins and viruses. As a topological surprise, a line of elegant blown-glass Klein bottles and their many-handled cousins grace a wall, not knowing their inside from their out.

Human touches are evident. Architect

Christopher Wren's brazen drafting dividers are suitably formidable for getting the measure of St Paul's cathedral in London. The skilled drawing of mathematical physicist Roger Penrose is clear in his book illustrations, which are shaded with a characteristic stippling technique to bring out the nuances of space and space-time. Crystallographer John D. Bernal's notebook sketches of crystal symmetries bring to life his search for the structure of the very graphite atoms his pencil left on the page, and the comfort of Tom Dixon's Pylon chair exploits the unique rigidity of triangles over other polygons.

Not all of the geometrical structures on display are of human construction. The University of Cambridge's Botanic Gardens have supplied fractal succulent plants that exhibit a range of self-similarity. Using dry ice, you can freeze your own snowflake and admire its six-fold symmetry, a large-

scale reflection of the geometrical symmetry of its molecular lattice. Look closely and you will see tiny deviations from that symmetry along the arms, each one unique because of slight differences in each snowflake's thermal history as it froze.

This vivid exhibition lays bare the many places where symmetry has a role in the world, and the importance of visualization in science and mathematics. Traditionally, pictures and models have been regarded as an unnecessary dilution of the true task of formal mathematics — to provide generalized proofs that make no appeal to particular examples. But to see these two activities in

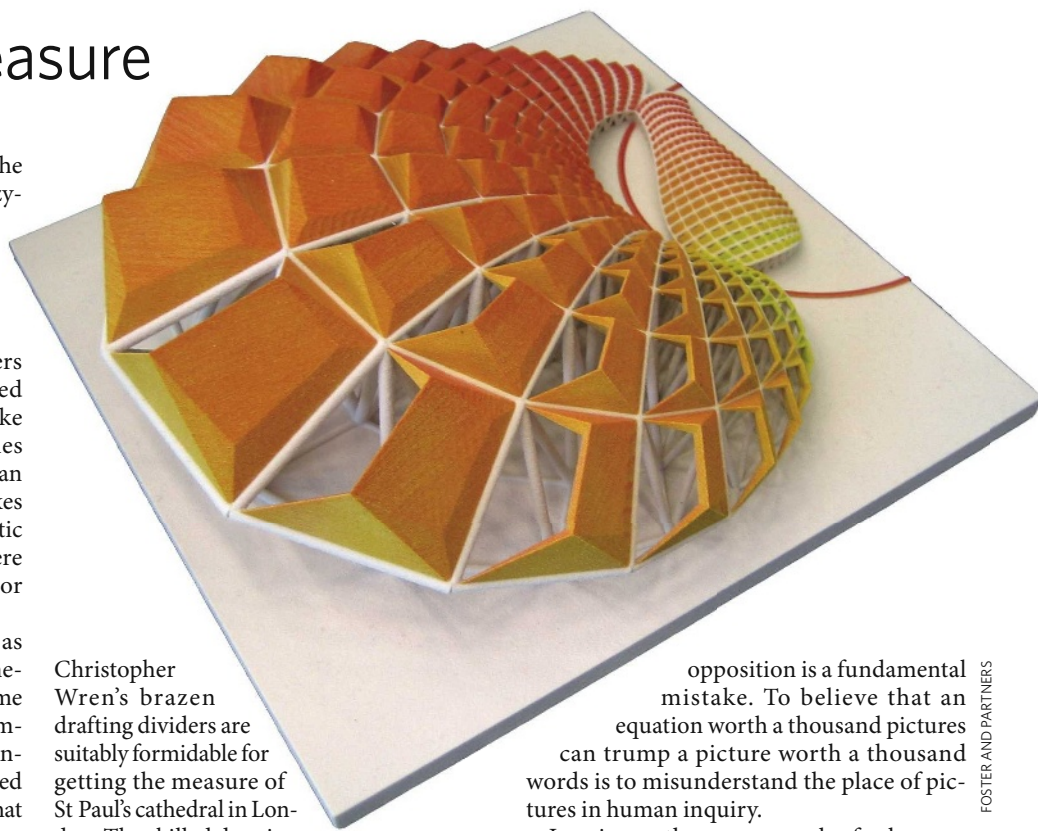
opposition is a fundamental mistake. To believe that an equation worth a thousand pictures can trump a picture worth a thousand words is to misunderstand the place of pictures in human inquiry.

In science, there are no rules for how we come by our theories. We can dream them up, see them in our tea leaves or take them out of the library. It doesn't matter. What does matter is that there is a rigorous process for testing those ideas to destruction through experiment and observation. Analogously, in mathematics there are no rules about how we gain intuitions about what might be true. Pictures can play an important part in that quest, but there is a formal process for establishing the truth or falsity of picturesque intuitions.

As in any small exhibition, other items could have been included. For example, no Vedic Sri Yantra or Islamic tessellated designs are displayed. The gallery's own geometry might have been contemplated, and more explanatory material would have been valuable, perhaps revealing why the Pylon chair needs to be made of triangles. Molecules such as the soccer-ball-shaped carbon-60 could have been shown, and at least one of the works of Maurits Escher, Oscar Reutersvärd or Lucio Saffaro deserve a place in any gallery of geometry. But these omissions suggest only that it would be good to have another exhibition like this. There is more to measure. ■

John D. Barrow is professor of mathematical sciences at the Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK, and author of *Cosmic Imagery: Key Images in the History of Science*.

***Beyond Measure: Conversations Across Art and Science* is at Kettle's Yard, Castle Street, Cambridge CB3 0AQ, UK, until 1 June ([www.kettlesyard.co.uk](http://www.kettlesyard.co.uk)).**



FOSTER AND PARTNERS

**"Displayed objects combine aesthetics with functionality; they were created to illustrate, to teach, to decorate or to demonstrate."**

## EXHIBITION

## Changing expressions

## Laura Spinney

From his early forties, the German artist Lovis Corinth developed a habit of painting a self-portrait every year, just before his birthday. In 1911, at the age of 52, he painted himself in classical style, as a knight in armour. A year later, he portrayed himself as Samson blinded, a picture of agony in chains and a loincloth. In between, the artist had suffered a stroke in the right hemisphere of his brain.

Corinth's style evolved so dramatically over his career, from naturalism to impressionism to expressionism, that he defies classification and is often considered an outsider. Art historians and neurologists debate whether the neurological damage he sustained drove him to be unconventional, or whether it was incidental to his artistic development.

Once the most fashionable portrait painter in Berlin, Corinth tired of the city after experiencing his stroke and turned to rural landscapes. Like the impressionists, he revisited the same spot — his beloved Walchensee in southern Germany — and painted it at different times of the day and year. In the last decade of his life, his preferred medium was drypoint,

a kind of printmaking, and his style became more sensitive and expressive.

One theme throughout his work is a fascination for flesh and the body. His interest in the visceral took him to places other painters feared to go, such as the slaughterhouse. But whereas his early paintings celebrated nudity in biblical scenes, his later ones dwelt on more morbid aspects, including the skeleton and death. The earlier works were subversive takes on the old masters, but still naturalistic; the later ones, with their violent brushstrokes and free use of primary colours, were much more expressive.

To resolve the issue of his evolving style, some scholars have turned to the unbroken record of the self-portraits. Corinth's stroke left him with the inability to process visuospatial information in the left side of his visual field, but how long this deficiency lasted is itself a matter of dispute. It is striking that in the earlier self-portraits he tends to be turned towards the left; in the later ones, towards the right. In an oil painting from 1925, he is also turned rightwards, but a mirror reflects his gaze back towards the left. Perhaps that was a final attempt at compensation — he didn't live to see his next birthday. ■



KUNSTHAUS, ZÜRICH

Lovis Corinth's *Last Self-portrait* may reflect stroke damage to his brain's right side.

Laura Spinney is a science writer based in London and Paris, and author of the novel *The Quick*.

**Lovis Corinth (1858–1925): *Between Impressionism and Expressionism* is at the Musée d'Orsay, Paris, until 22 June.**

## EXHIBITION

## Etching the artist's mind

## Colin Martin

Artist Susan Aldworth's interest in neuroscience was triggered during an emergency cerebral angiogram in 1999, while observing her brain's structure on a monitor. "You are looking inside your head while thinking, seeing, feeling; your brain is working while you are looking inside it," she marvelled.

In 2005, as artist-in-residence at the Royal London Hospital, UK, Aldworth sketched works on location in hospital clinics and collaborated with consultant neuroradiologist Paul Butler and neuropsychologist Paul Broks.

*Scribing the Soul* is her personal exploration of how matter becomes mind. Now on display in Oxford, UK, the exhibition will move to galleries across the United Kingdom next month.

In a series of prints entitled *Brainscape*, Aldworth chose etching as a medium for exploring cognition because, like neurotransmission, it uses chemical reactions. Delicate surface effects were created by drawing on metal plates with marker pens and then dipping the plates quickly in acid. Ghostly lines flicker across the

surfaces of the etchings, printed from the blue-linked plates. Her fast etching process is a metaphor for rapidly firing cerebral neurons, and the resulting cognitive 'blueprints' capture the nanoseconds when "flesh thinks".



Cerebral art: Susan Aldworth's *Between a Thing and a Thought*.

The surfaces of 20 etched plates, displayed as a wall-mounted grid, simultaneously absorb and reflect light, alternately evoking dull and scintillating thoughts. In two kaleidoscopic films, images of Aldworth's brain obtained during a functional magnetic resonance imaging scan are incorporated as a sequence of rapidly changing frames.

Aldworth's latest series of etchings, *The Self is a Shadow Puppet*, dramatizes the convoluted topography of the brain, its neural network and blood vessels. In one, a pair of ghostly hands reaches out of the dark background towards a disembodied brain, straining to touch its intangible, mysterious consciousness.

In his introduction to the exhibition, Broks makes the argument that our understanding of consciousness benefits from artists and scientists looking at it collaboratively. "The self is a shadow puppet shaped by the firings of a hundred billion brain cells," he writes. "These are conceptual conundrums. Intractable to current science, they call for an artistic response." ■ Colin Martin is a writer based in London.

***Scribing the Soul* runs at Science Oxford, Oxford, UK, until 23 May; at Peninsula Arts in Plymouth, UK, from 7 June to 18 July; and at the Transition Gallery in London from 26 July to 17 August.**

S. ALDWORTH



# The Impressionists' bible

Ogden Rood's textbook explaining the science of optics and colour mixing triggered an unexpectedly vivid response from the radical artists of the 1880s.

## Martin Kemp

The French Impressionist painters referred to the book *Modern Chromatics: With Applications to Art and Industry* as their "bible, and ... [they] carry it under their arm". So reported US painter Roland Rood to his father — the book's author — the distinguished physicist, Ogden Rood.

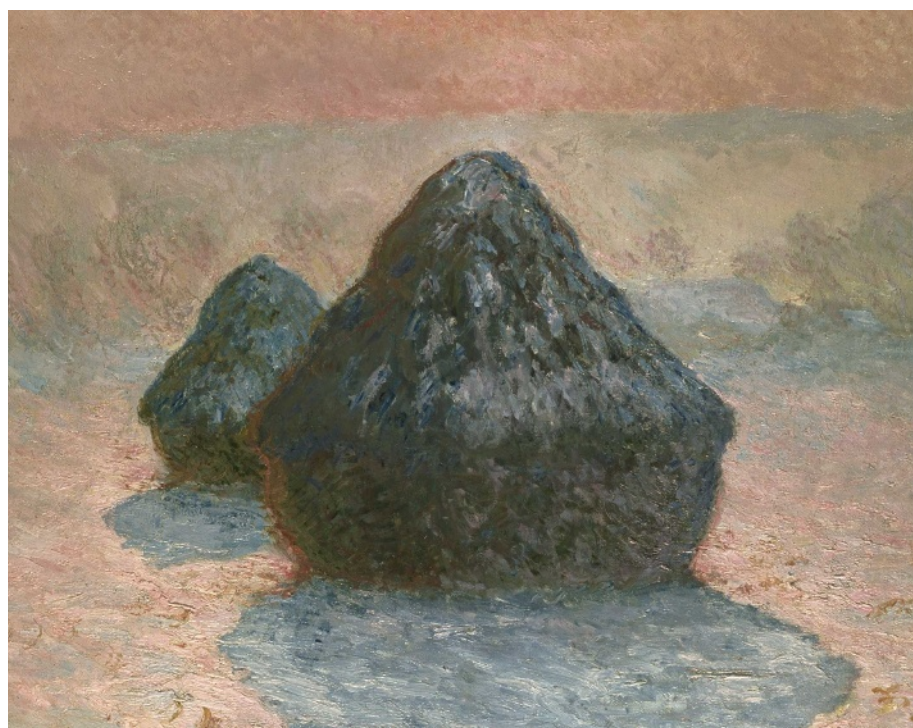
First published in 1879 by C. Kegan Paul and Company and translated into French as *Théorie Scientifique des Couleurs* in 1881, Ogden Rood's admired book intended to educate non-specialists in the latest optical research; in particular, the work of Hermann von Helmholtz on colour mixing. Ogden was himself an accomplished painter of watercolours, and saw it as his mission to teach artists about the science of colour.

According to correspondence collected in a 1906 article in *The Scrip* magazine, entitled Professor Rood's Theories of Color and Impressionism, Roland explained to his father that painters were paying special attention to "the oversensitivity of certain nerves of the eye to strong light, causing it to appear yellow; the dullness of certain nerves to weak light, making it bluish or purplish in tone; the principles of successive contrast, and more particularly of simultaneous contrast." He also explained their interest in the chapters on colour constants, on the duration of the impression on the retina, on colour mixture and on complementary colours.

Of these aspects of the perception of colour, Ogden's discussion of colour mixtures was perhaps most crucial to the Impressionist painters and their followers. Von Helmholtz established definitively the key distinction between the additive primaries (blue, red and green) for coloured lights and the subtractive primaries (blue, red and yellow) for pigment mixtures. Artists' theories on colour had previously failed to comprehend that light and pigments needed to be combined differently.

Ogden extended the experiments of physicist James Clerk Maxwell, who used spinning discs that mixed colours in different proportions. The disc mixtures were not additive because their variously coloured segments relied on the selective absorption of rays with different refractive properties. But they did suggest that painters might exploit optical mixtures on canvas rather than simply blending pigments on their palettes.

Ogden explained how different colours can be placed side by side — as lines or dots — and



Claude Monet used Ogden Rood's techniques to produce subtle colour mixes in *Haystacks: Snow Effect*.

then viewed at a distance such that the blending is more or less accomplished by the eye of the beholder. Under these circumstances, the tints mix on the retina and produce new colours that are identical to those obtained by the method of spinning discs.

The divided brushstrokes of Impressionism preceded the publication of Ogden's book, but in the later 1880s it became fashionable to seek a more scientific method of painting based on 'décompositions prismatiques' (the splitting of white light into the seven primaries identified by Isaac Newton). The divisionist or Pointilist painters, led by Georges Seurat, come most readily to mind. But some of the mainstream Impressionists also took note.

Claude Monet, in particular, began to show a more serious attitude to his analysis of colour. His famous series paintings, such as the *Haystacks* from 1891, reflect this new emphasis, systematically exploring the varied symphonies of colour that played across the same view at different times of day and during different seasons.

The painting that explores the 'snow effect' (one of more than 30 pictures of haystacks painted by Monet) blazes with complementary colours, above all orange and cyan blue. Colours

at the 'cold' end of the spectrum dominate. The complementaries were apparently applied in the studio, late in the execution of each piece and in a calculated manner.

At first sight, it seems that Ogden's impact on leading French painters met his fondest hopes. The reality was different, however. In their correspondence in *The Scrip*, Ogden responded gloomily to his son when asked about paintings by the "Frenchmen who call themselves Impressionists; some are by a fellow called Monet, others by a fellow called [Camille] Pissarro, and a lot of others".

"What do you think of them?", Roland ventured. "Awful! Awful!", Ogden gasped, horrified by the radical nature of the paintings.

When Roland told him what these painters said of his theories, Ogden became upset. He threw up his hands in horror and indignation, and cried: "If that is all I have done for art, I wish I had never written that book! My son, I always knew that a painter could see anything he wanted to in nature, but I never before knew that he could see anything he chose in a book."

As many educators have discovered, the results of arousing interest and propagating knowledge are not always what was expected. Monet's glorious *Haystacks* series demonstrates that this may not necessarily be a bad thing. ■ Martin Kemp is research professor in the history of art at the University of Oxford, OX1 1PT, UK.

**"The new emphasis explored the varied symphonies of colour that played across the same view at different times of day and during different seasons."**

## IN RETROSPECT

# When business became biology's plague

A 1920s best-seller about risky campus capitalism and early phage therapy still resonates today.

## Arrowsmith

by Sinclair Lewis

P. F. Collier: 1925. 448 pp.

## Thomas Häusler

Sinclair Lewis's novel *Arrowsmith* was a best-seller in 1925. It is a stinging satire of the US medical system from an era when quacks and their dubious cure-alls abounded — before the advent of antibiotics, double-blind trials and the Food and Drug Administration.

More than 80 years on, the problems of medical practice and research — as explored in the story of central character Martin Arrowsmith's capricious career from medical school to a top-notch research institute — are remarkably relevant. For example, as a young country doctor in North Dakota, Arrowsmith deals with the kind of militant anti-vaccinationists that recently caused child immunization rates for measles, mumps and rubella to plummet in the United Kingdom and in Switzerland, which is currently witnessing a measles epidemic. Indeed, one of the recurring themes of *Arrowsmith* is the difficulty that researchers have in communicating to the public.

The overarching thesis of this story is the powerful and sometimes damaging influence of business over medicine and science. During Arrowsmith's student days, most of his medical schoolmates are much more interested in their economic prospects than in learning. Later on in the book, he does a stint in the luxurious Rouncefield Clinic in Chicago, Illinois, where his bosses maximize their profits by arguing that: "Any proportion of the body without which people could conceivably get along should certainly be removed at once." Arrowsmith disdains this approach, and Lewis adds with a wink that: "The clinic did, perhaps, give over-many roentgenological examinations to socially dislocated women."

Lewis also satirizes the ease with which 1920s doctors and patients alike were apparently fooled, citing from a fictitious catalogue of a medical supplies company: "We guarantee that by the installation of a New Idea Panaceatic Electro-Therapeutic cabinet, you can increase your income from a thousand to ten thousand annually and please patients more than by the most painstaking plugging." This jibe brings to mind the numerous recent studies that show the large influence of even small gifts such as

pens on the prescribing habits of doctors, or that more expensive placebos have a greater healing effect than cheaper ones.

Lewis decries the effects of a rapacious business attitude on a scientific institution. At the climax of his career, Arrowsmith works in New York as a researcher at the privately funded McGurk Institute (modelled, scholars say, on New York's Rockefeller University). McGurk's director is a slick manager who sees his institution in a global competition for glory and cash. When Arrowsmith finds a cure for plague, his director pushes him to publish the results and dash into the clinic, even though there is insufficient

into Lewis's story along with his own ideals of medicine based on thorough science.

One incident in the book places these ideals in the foreground. Arrowsmith is summoned with his yet unproven plague medicine to a Caribbean island where an epidemic is raging. His mentor, called Max Gottlieb (modelled on de Kruif's own mentors Frederick Novy and Jacques Loeb), is adamant that Arrowsmith does a controlled study, giving his drug to only half of the sick. This was not common practice in the 1920s, and some characters in the novel denounce the act as inhuman — again, Lewis touches on a subject that is still hotly debated today.

The treatment that Arrowsmith uses in the study is a bacteriophage, a virus that kills bacteria — in this case, the plague-causing bacterium *Yersinia pestis*. At the time, this was cutting-edge science. Phages had been discovered less than ten years earlier and were used for treating certain infections for only about five years. When Arrowsmith's wife Leora, who accompanies him to the island, becomes infected with the plague and dies, her grieving husband abandons the trial protocol and treats everybody with phages. Again, Lewis's and de Kruif's analysis turned out to be an uncannily accurate reflection of the real-life history of phage studies.

Most actual phage trials were carried out without rigorous controls. Thus, a proper basis for the therapy was never laid — one reason why it was forgotten after the 1940s. Interestingly, infectious-disease specialists today are

once more studying phage therapy owing to widespread bacterial resistance to antibiotics. Unfortunately, the mistakes of the early days of phage research are still felt. Many researchers who are unfamiliar with the modern work still think of it as an obsolete method. If only the early phage therapists had read *Arrowsmith*.

To extricate himself from all the entanglements, Arrowsmith finally resigns from McGurk and lives with a colleague in the woods — like Henry Thoreau in *Walden*; or, *Life in the Woods* — where they make immune sera to finance their livelihood and their pure research. It is a utopian dream to which many pressurized researchers will still relate today.

Thomas Häusler is a science reporter for Swiss Public Radio DRS2 in Basel. He is author of *Viruses Vs Superbugs: A Solution to the Antibiotics Crisis?*



Scientific satire: Ronald Colman as Arrowsmith in the film of the same name.

evidence. The director is also unfazed by the necessity of fudging the data just a bit — the likes of Koch and Pasteur have to be beaten.

There were two reasons why Lewis was able to dissect the medical system so revealingly and why his analysis of its weaknesses is still largely valid. First, his father was a country doctor. Second, and more importantly, he had an informant: Paul de Kruif, the scientist-turned-science writer. De Kruif had worked on respiratory infections at what was then the Rockefeller Institute for Medical Research (now the Rockefeller University) and had an excellent insight into the guts of the medical system and its protagonists. He was fired from the institute after publishing a book chapter that was highly critical of US medicine. Its gist: too much business and mindless ritual, too little science. De Kruif instilled this sceptical view



# Lifting the fog from the north

A purist approach to wavefunctions can resolve some of quantum theory's infamous murkiness.

**Maximilian Schlosshauer**

As physicist Richard Feynman famously observed: "Nobody understands quantum mechanics." Whereas classical physics mirrors our everyday experiences — we can comprehend the idea of a tennis ball's velocity and apply Newton's laws of motion to explain it — this is not true of quantum physics. Classical concepts, such as position and momentum, fail to consistently describe phenomena of the atomic world.

In the early twentieth century, experiments showed that microscopic entities such as electrons sometimes behave more like waves than particles. Fired through slits, they form patterns similar to overlapping ripples in a pond. This prompted quantum mechanics' founders to introduce an abstract mathematical entity — the wavefunction — to describe physical systems.

Erwin Schrödinger in the 1920s thought that the wavefunction represented real entities. He suggested that when the wavefunction is spread out in space we see wave behaviour, and when it is constricted to a narrow 'wave packet' we get a particle. He hit a roadblock: quantum laws implied that a wave packet describing an isolated microscopic particle would rapidly disperse, yet real particles are obviously more stable.

The 'wavefunction only' picture was dismissed for other reasons. Evolving smoothly, the wavefunction seemed unable to account for processes that appear discontinuous in space and time. Quantum mechanics also predicted that wavefunctions would take strange shapes corresponding to a superposition of many classical-looking structures, such as Schrödinger's cat, doomed while hidden in its box to be in a state of being both alive and dead.

Physicists then reinterpreted the wavefunction as a symbolic device specifying the probabilities of measurement outcomes. This produced a dualism between the wavefunction — assumed to encapsulate everything about a system's physical state — and the random measurement outcomes, the quantum events. How do these events arise? And why can they be expressed in terms of classical variables?

Niels Bohr, one of the minds behind the Copenhagen interpretation of quantum mechanics, refused to go further. Thus descended the infamous 'fog from the north'. Classical measuring apparatuses took centre stage, precipitating the wavefunction's 'collapse' to give the experimental outcome. A dichotomy between quantum and classical was introduced. Incompatible classical



D. PARKINS

The classical world of our experience is a projected shadow in a quantum universe.

descriptions were sanctified using terminology such as wave-particle duality and complementarity (the doctrine that such descriptions must be associated with mutually exclusive experimental arrangements). Shadows of Copenhagen also hang over recent suggestions that quantum mechanics is fundamentally about information.

To lift the fog we should see the wavefunction as more than a tool for calculating probabilities. The wavefunction alone may suffice to understand both micro- and macroscopic phenomena once quantum mechanics is applied consistently.

A key ingredient is 'entanglement': when systems interact they lose their individuality and must be described by a shared wavefunction. Entanglement is ubiquitous. Physical systems cannot avoid interacting with their environment, so a system's behaviour is dictated by the wavefunction involving both system and environment. This is the physical process of quantum decoherence (see page 22). Decoherence censors the system so that only classical-looking structures are usually observable, such as objects in well-defined places. Observations, too, can be described as entangling interactions, without resorting to dualism.

Rapid yet continuous decoherence processes also explain quantum events. When an electron hits a photographic plate, it is not a mysterious act of 'measurement' that causes the electron's wavefunction to act as a point particle, but the entanglement between electron and plate (and its environment). The resulting decoherence establishes 'robustly narrow' wave packets that appear as particle-like entities, close to Schrödinger's idea.

The entangled wavefunction still involves a superposition of classical structures, such

as locations of spots on a photographic plate. Because we cannot fully control the environment (such as air molecules scattering off this page), such a superposition is usually hidden from our view. Still, why is a single spot here and not there? All such observations can be thought of as temporally extended and — thanks to decoherence — effectively autonomous 'branches' of the wavefunction. Arguably, this is all we need. Introducing some novel collapse process that rids the wavefunction of unobserved branches may offer some comfort but will risk disproof by experiment, although new physics may always be discovered.

Experimentalists have created many counterintuitive 'Schrödinger kittens' — including large molecules in a superposition of two distinct locations, and superpositions of electric current running in opposite directions. These have shown how decoherence leads to the continuous emergence of classical properties. Such observations can be explained using the 'pure quantum' picture of entangled wavefunctions, supporting the view that all classical structures are apparent. There is no need to introduce classical apparatuses, particles, quantum events, complementarity or wave-particle duality as fundamental concepts.

Taking the wavefunction seriously points to a quantum universe that is rather different from the everyday world. But as Feynman cautioned us, what may look like a paradox is often only a conflict between reality and your feeling of what reality ought to be. ■

**Maximilian Schlosshauer** is a research fellow at the School of Physics, University of Melbourne, Melbourne, Victoria 3010, Australia. He is author of *Decoherence and the Quantum-to-Classical Transition*.

See also pages 22 and 50.

ESSAY

## NEWS &amp; VIEWS

## NEUROSCIENCE

# Hidden female talent

Jai Y. Yu and Barry J. Dickson

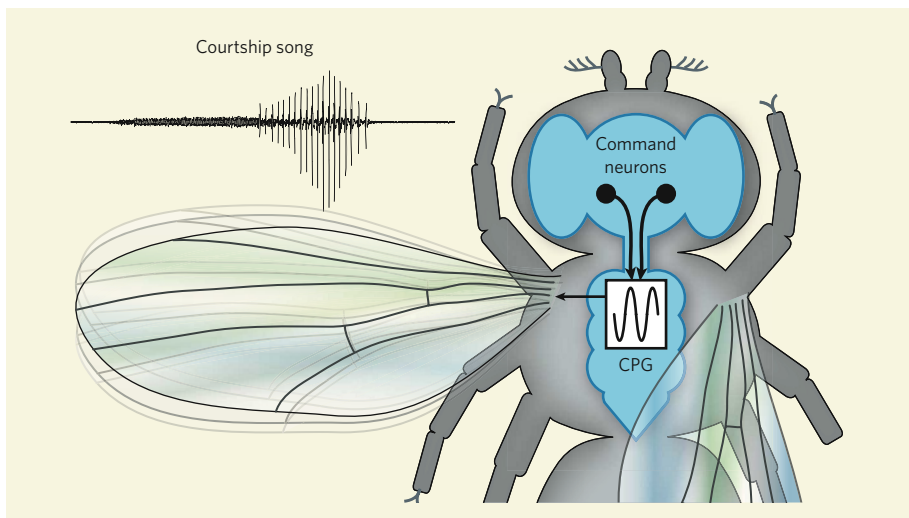
**A male fruitfly serenades his female with a courtship song produced by vibrating one wing. The female also has the neuronal circuitry to generate a song of her own, but her brain tells her not to.**

Male flies tend to put on quite a show when it comes to sex. They have to. Females can be rather choosy, expecting their suitors to demonstrate their quality through a demanding courtship ritual. In the fruitfly *Drosophila melanogaster*, as in many other insects, the male's most potent charm is his courtship song, which he produces by vibrating one wing. If his performance is sufficiently impressive, the female silently responds by allowing him to copulate with her. But females never sing themselves, suggesting that they lack the neural circuitry that produces these wing vibrations. Yet, in a remarkable paper published in the latest issue of *Cell*, Clyne and Miesenböck<sup>1</sup> show that the song circuitry is also present in the female thorax, but lies dormant. The essential difference, it seems, is all in the head.

The authors focused on a set of neurons that express the *fruitless* (*fru*) gene; this gene is the main sex-determination factor in the fly nervous system, at least when it comes to courtship<sup>2,3</sup>. It is expressed in some 2% of the fly's neurons, including central, sensory and motor neurons<sup>2,4,5</sup>. If synaptic communication by all these neurons is temporarily shut down, courtship behaviour, including singing, is selectively disrupted<sup>4,5</sup>. But is activation of these neurons sufficient to induce singing?

The trick Clyne and Miesenböck used to answer this question involves artificially stimulating neuronal activity with light, an approach pioneered by the Miesenböck group<sup>6,7</sup>. They first photoactivated all the *fru* neurons in intact and freely behaving males and found that nothing much happened. The reason for this lack of response could have been that simultaneous activation of all *fru* neurons results in conflicting excitatory and inhibitory signals. To confirm this, the authors took inspiration from the praying mantis and cut off the male's head. The female mantis figured out long ago that this drastic measure actually improves the male's sexual performance, possibly by eliminating inhibitory signals from the brain. Indeed, when the *fru* neurons were activated in decapitated male flies, the headless torso held out a wing and started to sing.

The Miesenböck group has previously shown<sup>6</sup> that photoactivation of a different set



**Figure 1 | Shining a light on courtship.** The male fruitfly produces a courtship song by vibrating one wing. Clyne and Miesenböck<sup>1</sup> show that these unilateral wing vibrations can be triggered in both males and females by using light to activate a central pattern generator (CPG) present in the thorax. The song CPG may normally be controlled by signals from descending 'command' neurons in the brain that differ between the sexes.

of neurons in the thorax can induce flies to beat both wings and fly away. Courtship song is characterized by slower and more complex vibrations that, in contrast to flight, are produced by just one wing. These wing vibrations are likely to be produced by neural circuits in the thorax that function as a central pattern generator (CPG). The CPG seems to include reciprocal inhibition, whereby the two sides of the nervous system inhibit each other to ensure that only one wing beats. Some of the *fru* neurons might feed signals into this CPG for song, or perhaps even be an intrinsic part of it (Fig. 1). There are some 650 *fru* neurons in the thorax, and so it will still take some effort to pin down the identity and function of the relevant ones.

So why don't female flies sing? By and large, *fru* neurons — including those in the thorax — are present in similar numbers and locations in both males and females<sup>4,5</sup>. Surprisingly, photoactivation of the *fru* neurons in the thorax of decapitated females, albeit more intense, also induces them to sing<sup>1</sup>. This observation indicates that females too have a functional song

circuit, even though they never use it. Why would females retain a circuit they never use? One possibility is that in female flies much of the song CPG overlaps with circuitries used for other behaviours. This has been shown<sup>8</sup>, for example, in molluscs, where distinct crawling and swimming behaviours are produced by CPGs with shared elements.

The female's song is not a perfect rendition of the male's. But how good is it? The best judge of that is another female fly, with her head intact. To investigate this, the authors used a ploy somewhat reminiscent of that of Edmond Rostand's hero in *Cyrano de Bergerac*, who wooed a fair damsel on behalf of his inarticulate friend. In Clyne and Miesenböck's assay, a pre-recorded song is played to a female as she is being courted by a male that cannot sing himself (because his wings have been cut off). The pre-recorded, light-triggered songs of decapitated males were effective seduction aids, but those of decapitated females were not. However, if the *fru* neurons of decapitated females were 'masculinized' by making them express the male-specific forms of *fru*, their



songs worked like a charm. So it seems that the male forms of *fru* fine-tune these neurons in the male to perfect his song.

Even if it is not well tuned, a song circuit is present in females. So what makes them hide their singing talent? The selective activation of thoracic song circuits in males but not females is likely to be controlled by some subset of the *fru* neurons in the brain. Indeed, classic studies of gynandromorph flies (which have a mixture of male and female nervous tissues) indicated<sup>4</sup> that certain brain regions must be 'male' to trigger the song. In this context, it is interesting to note that several pairs of neurons descending from the brain to the thorax are *fru*-positive. These neurons are prime candidates to convey sex-specific commands to the thoracic song circuits.

The picture that emerges from these studies is that the circuitry for song generation, like that for pheromone processing<sup>9,10</sup>, is largely shared between the sexes. The crucial sex differences seem to lie somewhere in between these bisexual input and output circuits, in dimorphic 'decision-making' centres in the brain. A similar design has recently been proposed<sup>11</sup> for the circuits that regulate sexual behaviour in mice: in females unable to perceive certain olfactory cues, male-like sexual behaviour results, presumably reflecting the activation of otherwise dormant circuits for these male behaviours in females. This modular and bisexual design affords considerable flexibility, which may even be exploited within the animal's own lifetime. Some species of fish, for example, change their sexual behaviour in response to social cues<sup>12</sup>. They may do this by simply resetting a few critical switches in the decision-making centres of an otherwise bisexual nervous system.

There is great excitement in neuroscience these days, as genetic tools are used to anatomically and functionally dissect the neural circuits that mediate complex animal behaviours<sup>13</sup>. Clyne and Miesenböck's work<sup>1</sup> beautifully illustrates the essential role photoactivation methods will have in this endeavour. As biochemists and biophysicists have long appreciated, surprising insights come when one can address questions of causality as well as correlation, reducing a system to its essentials and pushing it beyond its normal operating range. The mating behaviours of the humble fruitfly seem to be particularly amenable to this type of reductionist approach. ■

Jai Y. Yu and Barry J. Dickson are at the Research Institute of Molecular Pathology, Dr.-Bohr-Gasse 7, 1030 Vienna, Austria.  
e-mail: dickson@imp.ac.at

7. Zemelman, B. V., Lee, G. A., Ng, M. & Miesenböck, G. *Neuron* **33**, 15–22 (2002).
8. Popescu, I. R. & Frost, W. N. *J. Neurosci.* **22**, 1985–1993 (2002).
9. Kurtovic, A., Widmer, A. & Dickson, B. J. *Nature* **446**, 542–546 (2007).

10. Datta, S. R. *et al. Nature* **452**, 473–477 (2008).
11. Kimchi, T., Xu, J. & Dulac, C. *Nature* **448**, 1009–1014 (2007).
12. Shapiro, D. Y. *Science* **209**, 1136–1137 (1980).
13. Luo, L., Callaway, E. M. & Svoboda, K. *Neuron* **57**, 634–660 (2008).

## ELECTRONICS

# The fourth element

James M. Tour and Tao He

**Almost four decades since its existence was first proposed, a fourth basic circuit element joins the canonical three. The 'memristor' might herald a step-change in the march towards ever more powerful circuitry.**

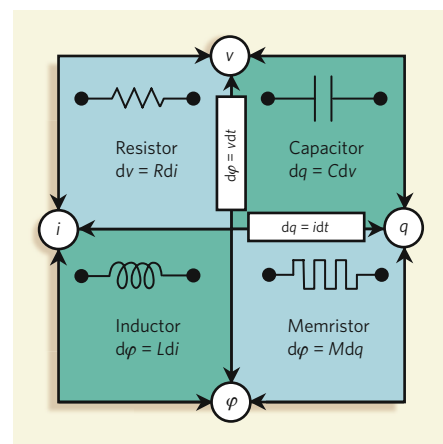
We learn at school that there are three fundamental two-terminal elements used for circuit building: resistors, capacitors and inductors. These are 'passive' elements, capable of dissipating or storing energy — but not, as active elements are, of generating it. The behaviour of each of these elements is described by a simple linear relationship between two of the four basic variables describing a circuit: current, voltage, charge and magnetic flux.

As the electrical engineer Leon Chua pointed out<sup>1</sup> in 1971, for the sake of the logical completeness of circuit theory, a fourth passive element should in fact be added to the list. He named this hypothetical element, linking flux and charge, the 'memristor' (Fig. 1). Almost 40 years later, Strukov *et al.*<sup>2</sup> (page 80 of this issue) present both a simple model system in which memristance should arise and a first, approximate physical example.

So what? Beyond its fundamental interest, the excitement lies in the possibility that the memristor could markedly extend how we can make electronic circuits work. In doing so, it might provide us with a way to keep on exponentially increasing computing power over time — thus maintaining something approximating to Moore's law, the rule-of-thumb to that effect that has been valid over the past few decades.

But before we get ahead of ourselves, some basics. According to the theory, a memristor is essentially a device that works under alternating current (a.c.) conditions<sup>1</sup> in which the applied voltage varies sinusoidally with time. As the polarity of this voltage changes, the memristor can switch reversibly between a less conductive OFF state and a more conductive ON state. Crucially, the value of the current flow through the memristor (the measure of its resistance) does not in the second half of the cycle retrace the exact path it took in the first. Because of this 'hysteresis' effect, the memristor acts as a nonlinear resistor the resistance of which depends on the history of the voltage across it — its name, a contraction of 'memory resistor', reflects just that property.

The memristor is a special case of a more



**Figure 1 | Complete quartet.** There are six independent permutations of two objects from a bank of four. Thus, six mathematical relations might be construed to connect pairs of the four fundamental circuit variables (current,  $i$ ; voltage,  $v$ ; charge,  $q$ ; magnetic flux,  $\phi$ ). Of these, five are well known. Two arise from the definitions of two of the variables concerned: charge and magnetic flux are the time integrals of current and voltage ( $dq = i dt$  and  $d\phi = v dt$ ), respectively. The other three lead to the axiomatic properties of three classic circuit elements: resistance,  $R$ , is the rate of change of voltage with current; capacitance,  $C$ , that of charge with voltage; and inductance,  $L$ , that of flux with current. The sixth relation leads to a fourth basic circuit element, which had been missing. Strukov *et al.*<sup>2</sup> have now found it: the memristor, with memristance,  $M$ , defined as the rate of change of flux with charge. (Figure adapted from refs 1 and 2.)

general class of nonlinear dynamical devices called memristive systems<sup>3</sup>. Whether physically realized or not, since memristance was first proposed the memristor has been successfully used as a conceptual tool for analysing signal processing and for modelling the workings of, for instance, electrochemical and nonlinear semiconductor devices.

Even so, the concept has not been widely adopted, possibly because in normal micro-scale chips the memristance is minute. But everything changes on the nanoscale, because

1. Clyne, J. D. & Miesenböck, G. *Cell* **133**, 354–363 (2008).
2. Baker, B. S., Taylor, B. J. & Hall, J. C. *Cell* **105**, 13–24 (2001).
3. Demir, E. & Dickson, B. J. *Cell* **121**, 785–794 (2005).
4. Manoli, D. S. *et al. Nature* **436**, 395–400 (2005).
5. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirián, L. & Dickson, B. J. *Cell* **121**, 795–807 (2005).
6. Lima, S. Q. & Miesenböck, G. *Cell* **121**, 141–152 (2005).

the size of memristance effects increases as the inverse square of device size. Strukov *et al.*<sup>2</sup> use a simple model to show how memristance arises naturally in a nanoscale system when electronic and atomic transport are coupled under an external voltage. The authors realize this memristive system by fabricating a layered platinum–titanium–oxide–platinum nanocell device. Here, the hysteretic current–voltage characteristics relate to the drift back and forth of oxygen vacancies in the titanium oxide layer driven by an applied voltage<sup>4</sup>.

This observation provides a wonderfully simple explanation for several puzzling phenomena in nanoscale electronics: current–voltage anomalies in switching; hysteretic conductance; multiple-state conductances (as opposed to the normal instance of just two conductance states, ON and OFF); the often mischaracterized ‘negative differential resistance’, in which current decreases as voltage increases in certain nanoscale two-terminal devices; and metal–oxide–semiconductor memory structures, in which switching is caused by the formation and breakdown of metal filaments owing to the movement of metal atoms under applied bias.

But what of Moore’s Law? Established by Intel co-founder Gordon Moore in 1965, this empirical rule states that the density of transistors on a silicon-based integrated circuit, and so the attainable computing power, doubles about every 18 months. It has held for more than 40 years, but there is a sobering consensus in the industry that the miniaturization process can continue for only another decade or so.

The memristor might provide a new path onwards and downwards to ever-greater processor density. By fabricating a cross-bar latch, consisting of one signal line crossed by two control lines<sup>5</sup>, using (two-terminal) memristors, the function of a (three-terminal) transistor can be achieved with different physics. The two-terminal device is likely to be smaller and more easily addressable than the three-terminal one, and more amenable to three-dimensional circuit architectures. That could make memristors useful for ultra-dense, non-volatile memory devices.

For memristor memory devices to become reality, and to be readily scaled downwards, the efficient and reliable design and fabrication of electrode contacts, interconnects and the active region of the memristor must be assured. In addition, because (unlike with transistors) signal gain is not possible with a memristor, work needs to be put into obtaining high resistance ratios between the ON and OFF states. In all these instances, a deeper understanding of the memristor’s dynamic nature is necessary.

It is often the simple ideas that stand the test of time. But even to consider an alternative to the transistor is anathema to many device engineers, and the memristor concept will have a steep slope to climb towards acceptance. Some

will undoubtedly trivialize the realization of this ubiquitous nanoscale concept, whereas others will embrace it only after the demonstration of a well-functioning, large-scale array of these densely packed devices. When that happens, the race towards smaller devices will proceed at full steam.

James M. Tour and Tao He are in the Departments of Chemistry, Computer Science, Mechanical Engineering and Materials Science, and the Smalley Institute for

Nanoscale Science and Technology, Rice University, Houston, Texas 77005, USA.  
e-mail: tour@rice.edu

1. Chua, L. O. *IEEE Trans. Circuit Theory* **18**, 507–519 (1971).
2. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. *Nature* **453**, 80–83 (2008).
3. Chua, L. O. & Kang, S. M. *Proc. IEEE* **64**, 209–223 (1976).
4. Yang, J. J. *et al.* *Nature Nanotech.* (in the press).
5. Kuekes, P. J., Stewart, D. R. & Williams, R. S. *J. Appl. Phys.* **97**, 034301 (2005).

## CLIMATE CHANGE

# Natural ups and downs

Richard Wood

**The effects of global warming over the coming decades will be modified by shorter-term climate variability. Finding ways to incorporate these variations will give us a better grip on what kind of climate change to expect.**

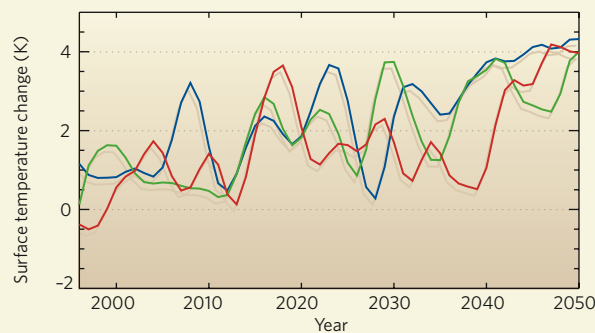
Climate change is often viewed as a phenomenon that will develop in the coming century. But its effects are already being seen, and the Intergovernmental Panel on Climate Change recently projected that, even in the next 20 years, the global climate will warm by around 0.2°C per decade for a range of plausible greenhouse-gas emission levels<sup>1</sup>. Many organizations charged with delivering water and energy resources or coastal management are starting to build that kind of warming into their planning for the coming decades. A confounding factor is that, on these timescales, and especially on the regional scales on which most planning decisions are made, warming will not be smooth; instead, it will be modulated by natural climate variations.

In this issue, Keenlyside *et al.* (page 84)<sup>2</sup> take a step towards reliably quantifying what those ups and downs are likely to be. Their starting point is the ocean. On a timescale of decades, this is where most of the ‘memory’ of the climate system for previous states resides. Anomalously warm or cool patches of ocean can be quite persistent, sometimes exchanging heat with the atmosphere only over several years. In addition, large ocean-current systems can move phenomenal amounts of heat around the world, and are believed to vary from decade to decade<sup>3,4</sup>.

To know and predict the state of the ocean requires an approach similar to

weather forecasting: one sets up (initializes) a mathematical model of the climate system using observations of the current state, and runs it forwards in time for the desired forecast period. With a given climate model, enough observations to set the ball rolling and a large-enough computer to move it onwards, the exercise is conceptually straightforward.

But does it actually produce anything useful? We don’t expect to be able to predict the details of the weather at a particular time several years in the future: that kind of predictability runs out after a week or two. But even predicting, say, that summers are likely to be unusually wet during the coming decade would be useful to many decision-makers. Only recently, with the study from Keenlyside *et al.*<sup>2</sup> and another



**Figure 1 | Heat up?** These three possible trends of winter temperature in northern Europe from 1996 to 2050 were simulated by a climate model using three different (but plausible) initial states<sup>6</sup>. The choice of initial state crucially affects how natural climate variations evolve on a timescale of decades. But as we zoom out to longer timescales, the warming trend from greenhouse gases begins to dominate, and the initial state becomes less important. Keenlyside and colleagues<sup>2</sup> use observations of the sea surface temperature to set the initial state of their model. Their results indicate that, over the coming decade, natural climate variability may counteract the underlying warming trend in some regions around the North Atlantic. (Figure courtesy of A. Pardaens, Met Office Hadley Centre).



from researchers at my own institution<sup>5</sup>, have climate modellers begun to explore whether such predictions are possible.

Keenlyside and colleagues' model<sup>2</sup> uses a very simple ocean initialization method in which they add heat to or remove it from the ocean surface until sea surface temperatures across the globe are close to observed values. They use their model to produce a set of retrospective 'forecasts' starting from earlier states, which they test against what actually happened. Their system produces refined temperature predictions a decade ahead for large parts of Europe and North America.

The enhanced accuracy of the model seems to stem from a greater ability to simulate natural variations in the meridional overturning circulation (MOC). This is a giant conveyor belt that brings warm water northwards into the North Atlantic, releases its heat to the atmosphere, and returns the cooled water to the south. There is evidence that the strength of this circulation can fluctuate naturally over periods of decades<sup>3</sup>; when it is strong, the climate in the North Atlantic region passes through a warm phase.

The authors use their model to predict that the MOC will weaken over the next decade, with a resultant cooling effect on climate around the North Atlantic. Such a cooling could temporarily offset the longer-term warming trend from increasing levels of greenhouse gases in the atmosphere. That emphasizes once again the need to consider climate variability and climate change together when making predictions over timescales of decades (Fig. 1).

These results provide encouragement that such predictions may be possible, but substantial points require clarification. Chief among these is whether the authors' initialization, which takes into account only sea surface temperatures, is in fact suitable to characterize the state of the MOC. The MOC extends to a depth of several kilometres and depends not just on temperature, but on the salinity of the ocean water. The answer will be some time in coming, as regular monitoring of the MOC has only just begun<sup>4</sup>.

In addition, there is the fact that, although it seems to improve predictions around the North Atlantic compared with models that have no initialization, the model's accuracy is less over other regions such as the tropical North Atlantic and central Africa. This might be because of deficiencies in the climate model or in the initialization procedure. The two studies that have attempted decadal-scale modelling to date<sup>2,5</sup> differ in the regions for which their predictions are most accurate, so clearly there is much still to be understood.

Keenlyside *et al.*<sup>2</sup> have so far used their model to predict decadal average temperature only. If it can be shown to accurately forecast other variables such as precipitation, as well as their variation during specific seasons, its range of applications would be greatly increased.

But climate predictions for a decade ahead will always be to some extent uncertain. There are inherent limits to the predictability of climate variability, and unpredictable perturbations to the climate, such as volcanic eruptions, cannot be factored in; predictions must inevitably be probabilistic in nature.

Even so, the first attempts at decadal prediction<sup>2,5</sup> suggest that reasonably accurate forecasts of the combined effects of increasing greenhouse-gas concentrations and natural climate variations can be made. Climate scientists are gearing up to test and extend these ideas over the coming years, in the hope that we can

then plan more confidently for the future. ■  
Richard Wood is at the Met Office Hadley Centre, FitzRoy Road, Exeter, Devon EX1 3PB, UK.  
e-mail: richard.wood@metoffice.gov.uk

1. Solomon, S. D. *et al.* (eds) in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).
2. Keenlyside, N. S., Latif, M., Jungclauss, J., Kornblueh, L. & Roeckner, E. *Nature* **453**, 84–88 (2008).
3. Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M. & Mann, M. E. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024233 (2005).
4. Cunningham, S. A. *et al.* *Science* **317**, 935–938 (2007).
5. Smith, D. M. *et al.* *Science* **317**, 796–799 (2007).
6. Stott, P. A. *et al.* *Science* **290**, 2133–2137 (2000).

## QUANTUM PHYSICS

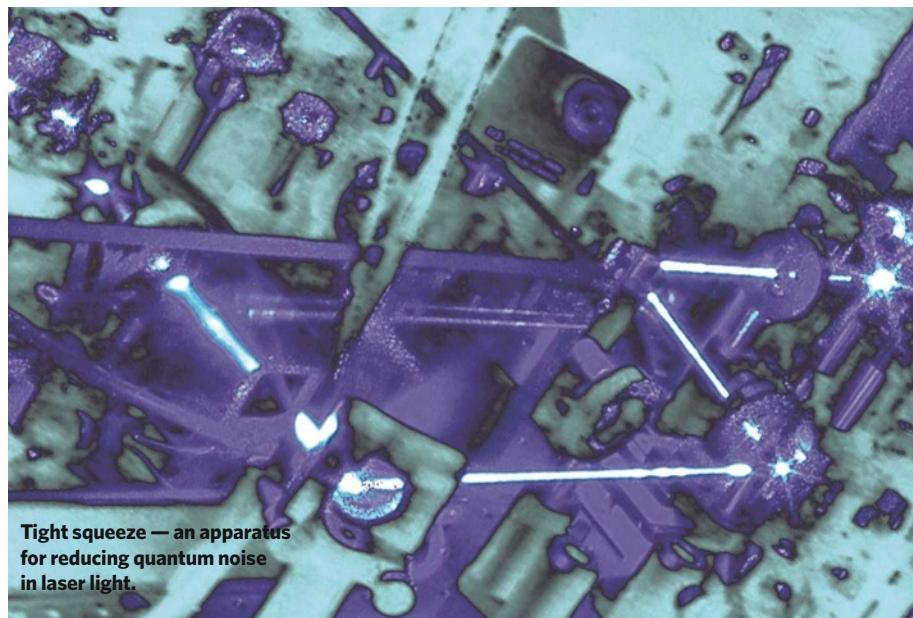
# The squeeze goes on

Eugene S. Polzik

**After 20 years of hard labour, squeezed states — light and matter whose quantum fluctuations have been arduously suppressed below standard levels of quantum noise — are coming of age and are ripe for application.**

Light is a quantum-mechanical object. One of the manifestations of that fact is that its amplitude (roughly, the number of photons present) and phase cannot be known simultaneously with arbitrary accuracy. Take, for example, an ideal laser pulse containing a mean number of photons  $N$ . The actual number of photons present in a pulse will obey poissonian statistics, with an uncertainty  $\sqrt{N}$ . Obeying the diktat of Heisenberg's uncertainty principle, the variance in the pulse's phase will accordingly vary as  $1/\sqrt{N}$ . These fluctuations form the characteristic 'shot noise' of light, and are the consequence of the fact that photons are statistically uncorrelated objects, fully independent of one another.

Because the Heisenberg principle puts the limit only on the product of the variances of amplitude and phase in a pulse of light, it is possible to reduce one or other of them beneath the shot-noise values. The catch is that the photons must be made correlated through some kind of interaction with matter. Three recent papers<sup>1–3</sup> present the latest advances in this technique of 'squeezing' light, and a fourth<sup>4</sup> tells us why it matters. Following a much earlier proposal<sup>5</sup>, this fourth paper describes a sensitivity-enhanced detector for gravitational waves — as-yet hypothetical tiny disturbances in space-time — that exploits squeezed-light states. This application is an illustration of how squeezed states have become central to



**Tight squeeze — an apparatus for reducing quantum noise in laser light.**

M. HÖST, NIELS BOHR INST.

quantum optics through their promise of accuracy improved beyond the standard quantum limit in interferometry<sup>6</sup>, measurements of atomic processes<sup>7</sup> and other sensing techniques.

Squeezed light was first generated in rubidium vapour<sup>8</sup> just over 20 years ago by exploiting correlations created by the interference of light of different frequencies, a technique known as four-wave mixing. That success was replicated shortly after in a nonlinear crystal<sup>9</sup> using parametric downconversion, which involves generating correlated photons by splitting a high-energy photon into two lower-energy ones. These first experiments achieved a reduction in the amplitude variance to 93% and 50% of the shot-noise values, respectively.

Spectacular as these experiments were as a proof of principle, the gain in signal-to-noise ratio fell some way short of that required for the technique to become useful in applications. Indeed, in most applications one can simply double the observation time and obtain the same improvement of sensitivity as produced with 50% squeezing. The tables are turned, however, as soon as light is squeezed down to around 10% of shot-noise variance. This is particularly true in applications such as gravitational-wave detection or the probing of sensitive atoms and molecules, where the exposure time for a measurement is limited.

Consequently, efforts towards this kind of extreme squeezing have been going on ever since the pioneering days. Now come some undeniable breakthroughs. In particular, Vahlbruch *et al.*<sup>1</sup> have used parametric downconversion, coupled with state-of-the-art materials and optical elements, to achieve 10 decibels of squeezing — equivalent to fluctuations at 10% of shot-noise variance. That squeezing increases to 5%, or 13 decibels, when corrected for detector inefficiency. Similarly, Takeno *et al.*<sup>2</sup> have achieved 9 decibels of squeezing. And in parallel, McCormick *et al.*<sup>3</sup> optimize the four-wave mixing method to squeeze the intensity difference between two optical beams by 8 decibels, corrected for detector inefficiency.

There is no reason to think that this is the end of the road. The degree of observed squeezing possible with parametric downconversion is limited primarily by optical losses and the limited quantum efficiency of photodetectors (that is, the probability that a photon will spark a measurable electrical signal); both will certainly be bettered in time.

Besides their use in ultra-precise measurements, squeezed states are also gaining relevance in quantum information science. They have been successfully employed for quantum teleportation<sup>10</sup>, and uses have also been proposed for them in global schemes for quantum computation and quantum communication<sup>11</sup>. In this context, it is particularly significant that the notion of squeezed states has of late propagated to atomic systems, where spin-squeezed states<sup>12</sup> have already been generated<sup>13</sup>. The uncertainties in the spin components of

an atomic ensemble in different directions are connected through a variant of Heisenberg's uncertainty principle, just as the uncertainties in light's phase and amplitude are. An ensemble is considered to be spin-squeezed if it has a mean spin  $J_x$  in a particular direction, but an uncertainty in the spin projection at right angles to that direction of less than  $\sqrt{(1/2) J_x}$ . Spin-squeezed states would allow standard quantum noise limits in atomic clocks and magnetometers to be overcome, and could also have their uses for the teleportation of atomic systems and for memory stores in quantum communication and quantum computing<sup>13</sup>.

Squeezed states were among the first non-classical states generated in the laboratory. As the number of applications hitting the limit to their sensitivity set by quantum noise grows, and as more squeezing becomes possible, the breadth of potential of squeezed states can only continue to expand.

Eugene S. Polzik is at QUANTOP — the Danish Quantum Optics Center, The Niels Bohr Institute,

University of Copenhagen, Blegdamsvej 17, 2100 København Ø, Denmark.  
e-mail: polzik@nbi.dk

1. Vahlbruch, H. *et al.* *Phys. Rev. Lett.* **100**, 033602 (2007).
2. Takeno, Y., Yukawa, M., Yonezawa, H. & Furusawa, A. *Opt. Exp.* **15**, 4321–4327 (2007).
3. McCormick, C. F., Boyer, V., Arimondo, E. & Lett, P. D. *Opt. Lett.* **32**, 178–180 (2007).
4. Goda, K. *et al.* *Nature Phys.* doi:10.1038/nphys920 (2008).
5. Caves, C. M., Thorne, K. P., Drever, R. W. P., Sandberg, V. D. & Zimmermann, M. *Rev. Mod. Phys.* **52**, 341–392 (1980).
6. Kimble, H. J. & Walls, D. (eds) *J. Opt. Soc. Am.* **B4** (10), (1987).
7. Polzik, E. S., Carri, J. & Kimble, H. J. *Phys. Rev. Lett.* **68**, 3020–3023 (1992).
8. Slusher, R. E. *et al.* *Phys. Rev. Lett.* **55**, 2409–2412 (1985).
9. Wu, L.-A., Kimble, H. J., Hall, J. L. & Wu, H. *Phys. Rev. Lett.* **57**, 2520–2523 (1986).
10. Furusawa, A. *et al.* *Science* **282**, 706–709 (1998).
11. Furusawa, A. *Phys. Rep.* **443**, 97–119 (2007).
12. Kitagawa, M. & Ueda, M. *Phys. Rev. A* **47**, 5138–5143 (1993).
13. Sherson, J., Julsgaard, B. & Polzik, E. S. *Adv. Atom. Mol. Opt. Phys.* **54**, 81–130 (2006).

## IMMUNOLOGY

# T cells hang in the balance

Emily A. Stevens and Christopher A. Bradfield

**Equally important as the immune system's function in fighting invaders is its ability to tolerate self. But environmental toxins could shift the equilibrium between these activities one way or the other.**

To ensure the efficient execution of its various activities, the immune system has distributed responsibility between different sets of its cells. But disease, pathogens and pollutants might disrupt the levels and functions of these cells. Take two subsets of T cells for instance:  $T_{reg}$  cells are involved in recognizing and 'tolerating' the body's own cells and molecules as harmless;  $T_H17$  cells, by contrast, promote inflammation. In this issue, Quintana *et al.*<sup>1</sup> (page 65) and Veldhoen *et al.*<sup>2</sup> (page 106) demonstrate that environmental toxins tilt the balance in the levels of these two cell types, altering the severity of experimental autoimmune encephalitis (EAE), a condition in mice that mimics multiple sclerosis.

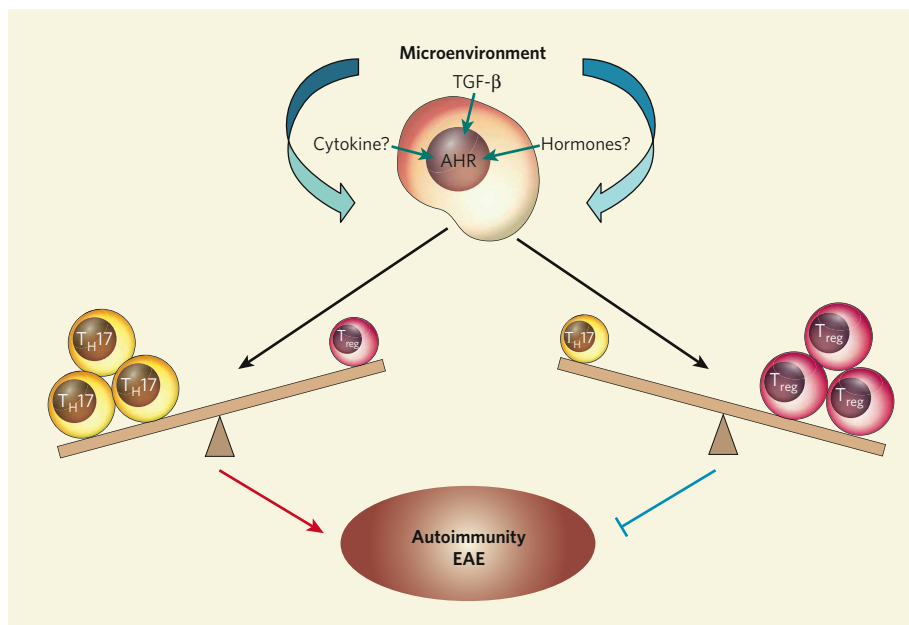
Control of immune responses is essential to health.  $T_{reg}$  cells suppress both the proliferation of effector cells and the secretion of immune mediator proteins called cytokines. The outcome can be positive (a reduction in autoimmune disease and allergy) or negative (limited immune response to infections or inhibition of anti-tumour responses)<sup>3</sup>. Similarly, by promoting inflammation,  $T_H17$  cells can either participate in the clearance of external pathogens or cause autoimmunity<sup>4</sup>. The cytokine microenvironment of activated T cells seems to influence whether they differentiate into  $T_{reg}$  or  $T_H17$  cells. For example, adding the cytokine

TGF- $\beta$  to T cells in culture induces their differentiation into  $T_{reg}$ , whereas in the presence of another cytokine, IL-6, TGF- $\beta$  promotes production of  $T_H17$  cells<sup>3,4</sup>.

Both  $T_{reg}$  and  $T_H17$  cells express the aryl hydrocarbon receptor (AHR) — a ligand-activated gene transcription factor that has pathological as well as physiological activity. In mammals, the AHR is well known as a mediator of the toxicity of environmental pollutants, including its prototype ligand, dioxin. The toxic effects of dioxin include thymic atrophy, chloracne, tumour promotion and wasting, and can even lead to death. In addition, studies of AHR-deficient mice have uncovered numerous physiological roles for this receptor, although the ligands that mediate these roles are yet to be discovered. Considerable experimental evidence indicates<sup>5,6</sup> that the AHR mediates its physiological roles, as well as most, if not all, of the toxic effects of dioxin through alterations in gene expression.

Quintana *et al.*<sup>1</sup> and Veldhoen *et al.*<sup>2</sup> now establish a link between the AHR and  $T_{reg}/T_H17$  differentiation.  $T_H17$  cells secrete the pro-inflammatory cytokines IL-17 and IL-22, which are required for effective clearance of some pathogens<sup>4</sup>. Veldhoen *et al.* show that activation of the AHR increases the expression of these cytokines in cultured  $T_H17$  cells,





**Figure 1 | One cell's poison is another cell's antidote.** Regulatory T cells ( $T_{reg}$ ) suppress the immune system, whereas  $T_H17$  cells promote inflammation. Veldhoen *et al.*<sup>2</sup> demonstrate that activation of the transcription factor AHR in  $T_H17$  cells increases expression of pro-inflammatory cytokines and worsens experimental autoimmune encephalitis (EAE). Quintana *et al.*<sup>1</sup> show that AHR signalling in  $T_{reg}$  cells increases their activity and dampens EAE. TGF- $\beta$  is involved in both  $T_{reg}$  and  $T_H17$  cell differentiation. Through its role as an environmental sensor, AHR might ensure an equilibrium between these two T-cell subpopulations during an immune response via its interactions with the TGF- $\beta$ -mediated signalling pathway.

which would be consistent with increased  $T_H17$ -mediated inflammation *in vivo*. They also find that, on induction of EAE in AHR-deficient mice, the absolute number of  $T_H17$  cells is reduced, whereas the number of  $T_{reg}$  cells remains unchanged.

The hallmark of  $T_{reg}$  cells is expression of the transcription factor Foxp3, which is required for the suppressive activity of these cells. Quintana *et al.*<sup>1</sup> show that the AHR directly regulates Foxp3 expression. Moreover, they demonstrate that whether the AHR shifts the balance in favour of  $T_{reg}$  cells or  $T_H17$  cells depends on the ligand that activates it. Dioxin increases  $T_{reg}$  activity and proliferation, decreases the number and function of  $T_H17$  cells, and suppresses EAE. Another potent activator of the AHR, 6-formylindolo[3,2-b]carbazole (FICZ), has the opposite effect<sup>1,2</sup>: it increases  $T_H17$ -cell activity and exacerbates EAE.

The AHR is a member of the PAS family of transcription factors, which are known as environmental sensors<sup>7</sup>. Being a transcription factor, the AHR is poised to fine-tune signalling at the level of gene expression. It can therefore probably sense and integrate environmental cues, such as cytokines, hormones and chemicals, as well as modulate the immune response by affecting  $T_H17/T_{reg}$  cell differentiation.

The TGF- $\beta$ -mediated signalling pathway is also involved in both  $T_H17$  and  $T_{reg}$  cell differentiation<sup>1,2</sup>, and interactions between the AHR and TGF- $\beta$  signalling pathways have been characterized in many contexts<sup>8</sup>. Furthermore, Quintana and colleagues' results<sup>1</sup>

indicate that dioxin influences  $T_{reg}$  differentiation through TGF- $\beta$ . They show that TGF- $\beta$  mimics dioxin's effects on  $T_{reg}$  cells and that inhibiting TGF- $\beta$  signalling suppresses dioxin-induced  $T_{reg}$  activity. FICZ also seems to modulate TGF- $\beta$  activity<sup>1</sup>. So it is by modulating TGF- $\beta$  signalling within the nucleus that the AHR is likely to shift the balance between

the two T-cell populations with opposing effects (Fig. 1).

Although a physiological role for the AHR in regulating the levels of  $T_{reg}/T_H17$  cells would be intriguing, the pharmacology of this system is far from clear. A reason for the conflicting effects of dioxin and FICZ on EAE could be the pharmacology or pharmacokinetics of these chemicals. The AHR solely mediates the effects of dioxin, whereas FICZ might affect additional signalling pathways. Also, whereas FICZ is rapidly metabolized, dioxin is not. Nonetheless, the degree or timing of AHR stimulation with these chemicals could mimic various micro-environmental cues that a developing T cell might receive from its natural environment. Understanding how specific AHR ligands lead to different outcomes *in vivo* will not only provide information about AHR biology, but will also shed light on how the levels of  $T_{reg}$  and  $T_H17$  cells regulate the immune response. This knowledge is crucial for any potential therapeutic approach directed at the AHR. ■

Emily A. Stevens and Christopher A. Bradfield are in the McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706, USA.

e-mail: bradfield@oncology.wisc.edu

1. Quintana, F. J. *et al.* *Nature* **453**, 65–71 (2008).
2. Veldhoen, M. *et al.* *Nature* **453**, 106–109 (2008).
3. Mottet, C. & Golshayan, D. *Swiss Med. Wkly* **137**, 625–634 (2007).
4. LaRosa, D. F. & Orange, J. S. *J. Allergy Clin. Immunol.* **121**, S364–S369 (2008).
5. Bunger, M. K. *et al.* *J. Biol. Chem.* **278**, 17767–17774 (2003).
6. Walliser, J. A. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 16677–16682 (2004).
7. Gu, Y. Z., Hogenesch, J. B. & Bradfield, C. A. *Annu. Rev. Pharmacol. Toxicol.* **40**, 519–561 (2000).
8. Chang, X. *et al.* *Mol. Cell. Biol.* **27**, 6127–6139 (2007).

## NETWORKS

# Teasing out the missing links

Sid Redner

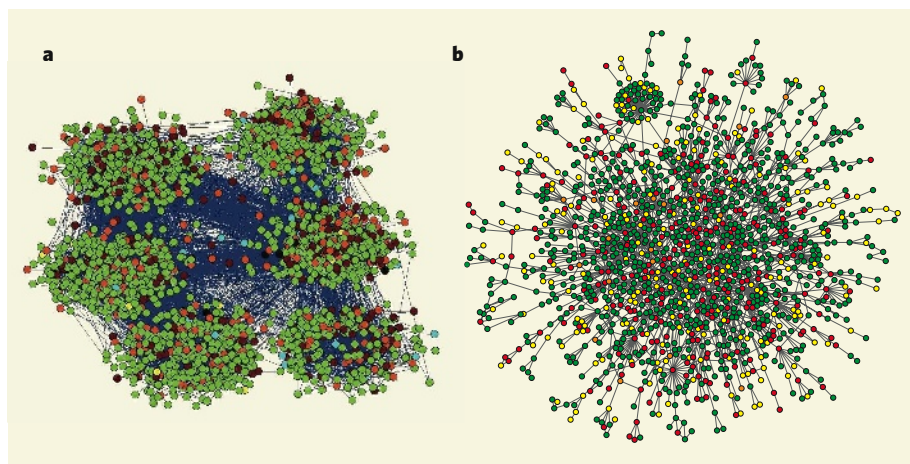
**Focusing on the hierarchical structure inherent in social and biological networks might provide a smart way to find missing connections that are not revealed in the raw data — which could be useful in a range of contexts.**

As human beings, we are all participants in complex, interlocking social networks<sup>1</sup>. As the information revolution gathers pace, the scope and reach of those networks is rapidly expanding. The World Wide Web provides easy connections to informational, commercial and recreational websites. Many people, especially the young, are hooked up to social-networking websites. Social bookmarking, in which participants share links to their favoured websites, is the latest craze.

In this increasingly tangled web, is it possible to make sense of the patterns of connections

between people, and so perhaps learn something useful? In this issue, Clauset, Moore and Newman<sup>2</sup> (page 98) introduce an appealing, simple and flexible model to do just that: the 'hierarchical random graph'.

Their starting point is the well-known hierarchical structure of a family tree, or dendrogram. We are genetically connected to our siblings (our 'zeroth cousins') through our parents, to our first cousins by our grandparents, to our second cousins through our great-grandparents, and so on, onwards and upwards. What is the probability that we actually know



**Figure 1 | Assortative, disassortative.** **a**, In assortative networks, well-connected nodes tend to join to other well-connected nodes, as in many social networks — here illustrated by friendship links in a school in the United States<sup>6</sup>. **b**, In disassortative networks, by contrast, well-connected nodes join to a much larger number of less-well-connected nodes. This is typical of biological networks; depicted here is the web of interactions between proteins in brewer's yeast, *Saccharomyces cerevisiae*<sup>7</sup>. Clauset and colleagues' hierarchical random graphs<sup>2</sup> provide an easy way to categorize such networks. (Part **a** reproduced with permission from ref. 6.)

our cousins of the *n*th degree? In most cases, it is obvious that this probability decreases as the degree of separation in the hierarchy increases: we are more likely to know close family members than distant relatives.

This basic family tree describes a highly unrealistic, insular population: only children and parents know each other directly (are directly connected in the tree), and no other social connections exist. Clauset *et al.*<sup>2</sup> incorporate a more social and random aspect into their model by adding lateral connections between nodes in the tree that share a common ancestor. The probabilities of these connections depend in a general way on the closeness of the nodes in the tree, and can be adjusted to describe a diverse range of networks in the social and biological worlds. The flexibility afforded by their model provides a simple way to interpolate between the signature network behaviours of assortativity and disassortativity<sup>3,4</sup> (Fig. 1).

The model can thus account for the myriad ways in which social participants can mix. One can, in fact, identify reversed patterns of sociability in which connection increases with degree of separation: that is, where distant relatives are more likely to know each other than be acquainted with members of their own nuclear family. This apparently bizarre situation seems to arise in food webs<sup>5</sup>, in which two predator species may prey on the same species without preying on (being acquainted with) each other, and species that are many feeding levels apart, and thus only distantly related, routinely eat or are eaten by each other.

Perhaps the most intriguing application of the hierarchical random graph construction is the possibility of efficiently predicting missing links in networks in which the available information is incomplete. Naively, one should query every pair of ostensibly unconnected nodes

to uncover all the missing links. But such an exhaustive search is grossly inefficient, because the computational effort grows as the square of the number of nodes. By first taking all known network connections and statistically fitting them to a hierarchical random graph, one can infer the dependence of the lateral-connection probabilities on the depth of the nodes in the hierarchy. One then restricts queries to node pairs for which the probability exceeds a specified threshold, a computationally much more efficient process.

The authors tested the effectiveness of their

method by removing connections at random from three sets of network data: a bacterial metabolic network; a food-web among grass-land species; and, most piquantly, the network of associations among terrorist cells. Next, they determine the lateral-connection probabilities of the underlying hierarchical random graph that best represents the incomplete network. Finally, they test for missing links by looking for unconnected node pairs with a high lateral-connection probability. In a related vein, the method can also ferret out false-positive links: links that appear in the data, but exist between nodes with a low lateral-connection probability.

In many cases, the new method was more reliable in reconstructing the true network structure than were other commonly used algorithms. In our ever-more-interconnected world, there is an increasing need for such theoretical tools that provide a more intimate understanding of the connectivity of complex networks.

Sid Redner is in the Center for Polymer Studies and the Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA.  
e-mail: redner@bu.edu

1. Wasserman, S. & Faust, K. *Social Network Analysis* (Cambridge Univ. Press, 1994).
2. Clauset, A., Moore, C. & Newman, M. E. J. *Nature* **453**, 98–101 (2008).
3. Newman, M. E. J. *Phys. Rev. Lett.* **89**, 208701 (2002).
4. Newman, M. E. J. *Phys. Rev. E* **67**, 026126 (2003).
5. Dawah, H. A., Hawkins, B. A. & Claridge, M. F. *J. Anim. Ecol.* **64**, 708–720 (1995).
6. González, M. C., Herrmann, H. J., Kertész, J. & Vicsek, T. *Physica A* **379**, 307–316 (2007).
7. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. *Nature* **411**, 41–42 (2001).

## ASTROPHYSICS

# Rays from the dark

Rainer Plaga

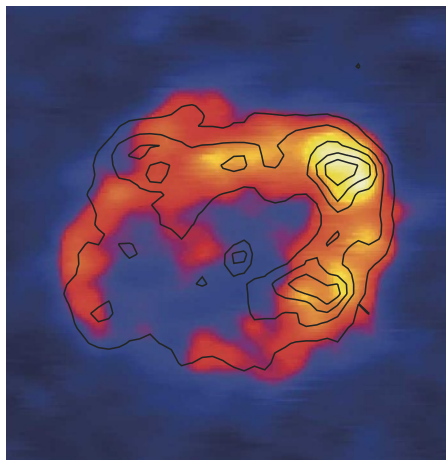
**The origin of the cosmic rays that bombard Earth has troubled physicists for nigh on a century. Supernova remnants are a favoured source — but we should keep our minds open to alternatives.**

Cosmic rays fall in a steady, imperceptible rain onto Earth. Despite their stealth, they are thought to influence both the climate of our planet and the evolution of its inhabitants. They are mainly protons and heavier nuclei, from helium upwards, and come in a wide range of energies. The most energetic are now thought to come from the active nuclei of remote galaxies<sup>1</sup>. But the origin of by far the larger component of the cosmic rain — 'galactic' cosmic rays of lower energies, from around a gigaelectronvolt to more than a million times that, 3 petaelectronvolts (PeV) — is an entirely different conundrum.

One thing we do know is that galactic cosmic

rays follow an unbroken power law, decreasing steadily in their frequency of occurrence up to the 'knee energy' of 3 PeV. Beyond that point, the downward slope abruptly steepens. This striking unity in the data invites the conclusion that galactic rays come from one type of source. But what is that source? Last year, Uchiyama *et al.*<sup>2</sup> reported unexpected X-ray observations that seemed to pinpoint it: galactic cosmic rays are produced when a star explodes, creating a supernova that violently expels a proportion of its matter into interstellar space to form a 'supernova remnant'. The propagation of this stellar debris drives a shockwave that, aided by preexisting compressed magnetic fields,





**Figure 1 | Promising candidate.** The shell of the supernova remnant SNR RX J1713.7–3946, a candidate source of cosmic rays<sup>2</sup>, is seen here in the light of very-high-energy  $\gamma$ -rays (intensity colour-coding) and X-rays (black contours). Whether either or both of these emissions provide clinching evidence that the supernova remnant is a typical source of cosmic rays is still the subject of debate. (Figure reproduced from ref. 10.)

can accelerate protons and nuclei of the interstellar matter. This accelerating kick can be repeated many times if a particle happens to be deflected back into the medium in front of the shockwave.

This process is known as diffusive shock acceleration, and could account both for the existence of the galactic cosmic rays recorded on Earth and for their wide distribution in energy. But there is a problem: standard supernova shocks cannot accelerate protons to energies beyond 0.1 PeV, a gaping factor of 30 below the knee energy<sup>3</sup>. To bridge this gap, a mechanism for the more efficient conversion of shockwave energy to magnetic-field energy has been proposed<sup>4</sup>. The amplified magnetic fields thus generated can deflect particles more rapidly into the region in front of the shock, allowing more frequent accelerating kicks.

Uchiyama *et al.*<sup>2</sup> argued that they had found evidence of just such an amplified magnetic field. Between 2000 and 2006 they made repeated observations of the debris from a supernova that, according to records made by Chinese astronomers, had exploded in February or March of AD 393. This supernova remnant, known as SNR RX J1713.7–3946, has since grown to a diameter of about 60 light years, and radiates intensely at X-ray wavelengths (Fig. 1).

As the authors revealed, the remnant contains compact areas, smaller than a light year across, of enhanced X-ray luminosity. These hotspots flare up and down on a timescale of about a year. After considering other possible explanations for the flaring, Uchiyama *et al.* concluded that it must indicate rapid acceleration of electrons, followed by their cooling through X-ray emission. The cooling time

depends on the magnetic-field strength, and the observed speed of the fading indicated an amplified ambient magnetic field strong enough to accelerate charged particles to knee energies, and so account for the profile of cosmic-ray energies seen on Earth.

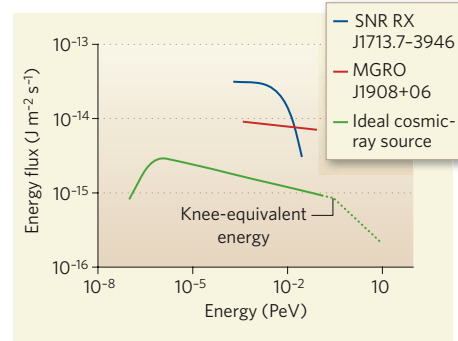
But caveats to this interpretation have emerged. First, the evidence stems exclusively from the behaviour of electrons; plausible models of magnetic-field amplification near hotspots suggest that the process is not sufficient to accelerate protons to knee energies<sup>5</sup>. Second, the possibility has been floated that the observed X-ray flares do not originate in electron cooling at all, but instead reflect modest variations in magnetic-field strength over short timescales that are unrelated to particle acceleration<sup>6</sup>.

Direct, ‘smoking gun’ evidence for the acceleration of protons to the requisite energies might come from very-high-energy (VHE)  $\gamma$ -rays and neutrinos, which would be created when protons and nuclei, freshly accelerated in a source, collide with ambient matter. This process creates the particles known as pions, which decay into electrons,  $\gamma$ -rays and neutrinos. The spectrum of  $\gamma$ -rays that would emerge if this process were occurring would follow a constant power-law extending from about 100 megaelectronvolts up to a ‘knee-equivalent’ energy of about 0.3 PeV.

SNR RX J1713.7–3946 is indeed a very luminous source of such VHE  $\gamma$ -rays, although it has recently been argued that this radiation is produced by energetic electrons, rather than by accelerated protons and nuclei<sup>7–9</sup>. Moreover, the spectrum nosedives and then cuts off at an energy of about 0.018 PeV, more than 10 times below the knee equivalent<sup>10</sup> (Fig. 2).

Recently, a sensitive and complete survey of VHE sources in the Northern Hemisphere’s sky was published<sup>11</sup>. Three of the five most luminous sources found are of a novel type, dubbed ‘dark accelerator’ because, in contrast to supernova remnants, no radiation from them has so far been identified in any other part of the electromagnetic spectrum. The energy distribution of one of them, labelled MGRO J1908+06, has been measured up to energies of about 0.09 PeV, and future observations with existing and projected ground-based experiments will find out if its spectrum in fact extends up to knee-equivalent energies.

Observations with the soon-to-be-launched GLAST observatory will reveal whether the low-energy spectrum of this source fits the bill for a typical cosmic-ray source. Further supporting evidence will come from several years of observations of cosmic neutrinos at the IceCube detector now under construction at the South Pole. If enough neutrinos are present, that would prove the acceleration of protons, as accelerated electrons can only produce a negligible number of neutrinos. But whatever the nature of the dark accelerators, they must be better understood before the last word on cosmic-ray origin can be spoken.



**Figure 2 | Return to the source.** The ideal candidate source for galactic cosmic rays would reproduce features of the cosmic-ray spectrum at lower energies in its  $\gamma$ -ray spectrum (here shown in a highly schematic form), with the  $\gamma$ -ray energy flux decreasing with a smooth power-law dependence up to an energy of around 0.3 PeV, corresponding to the ‘knee’ of observed cosmic rays. The  $\gamma$ -ray spectrum of SNR RX J1713.7–3946 drops off sharply well below the knee-equivalent energy; that of the ‘dark accelerator’ MGRO J1908+06 shows no such cut-off, but measurements do not yet extend to the necessary higher energies. The absolute normalization of the energy fluxes should not be compared owing to uncertainties in the distances of the sources.

Future research might confirm the conclusions that supernova remnants are the main source of cosmic rays. Alternatively, compelling observational evidence for some completely different source class — intriguing objects with dramatic-sounding names such as ‘superbubbles’, relativistic ‘fireball transients’ and ultrarelativistic ‘cannonballs’ — might be found. Whatever the answer, these are happy times for cosmic-ray physics: observers can be proud to have discovered cosmic objects completely unexpected by theory, and theorists can rejoice over a plethora of clues at their disposition to solve this outstanding problem of modern physics once and for all.

Rainer Plaga is in the Department for New Technologies and Scientific Foundations, Federal Office for Information Security (BSI), 53175 Bonn, Germany.

e-mail: rainer.plaga@bsi.bund.de

1. The Pierre Auger Collaboration *Science* **318**, 938–943 (2007).
2. Uchiyama, Y., Aharonian, F. A., Tanaka, T., Takahashi, T. & Maeda, Y. *Nature* **449**, 576–578 (2007).
3. Lagage, P. O. & Cesarsky, C. J. *Astron. Astrophys.* **125**, 249–257 (1983).
4. Bell, A. R. & Lucek, S. G. *Mon. Not. R. Astron. Soc.* **321**, 433–438 (2001).
5. Ellison, D. C. & Vladimirov, A. *Astrophys. J.* **673**, L47–L50 (2008).
6. Butt, Y., Porter, T., Katz, B. & Waxman, E. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/0801.4954> (2008).
7. Huang, C.-Y., Park, S. E., Pohl, M. & Daniels, C. D. *Astrophys. J.* **27**, 429–439 (2007).
8. Katz, B. & Waxman, E. *J. Cosmol. Astropart. Phys.* **01**, 018 (2008).
9. Plaga, R. *New Astron.* **23**, 73–76 (2008).
10. Aharonian, F. A. *et al. Astron. Astrophys.* **464**, 235–243 (2007).
11. Abdo, A. A. *et al. Astrophys. J.* **664**, L91–L94 (2007).

## OBITUARY

## John Wheeler (1911–2008)

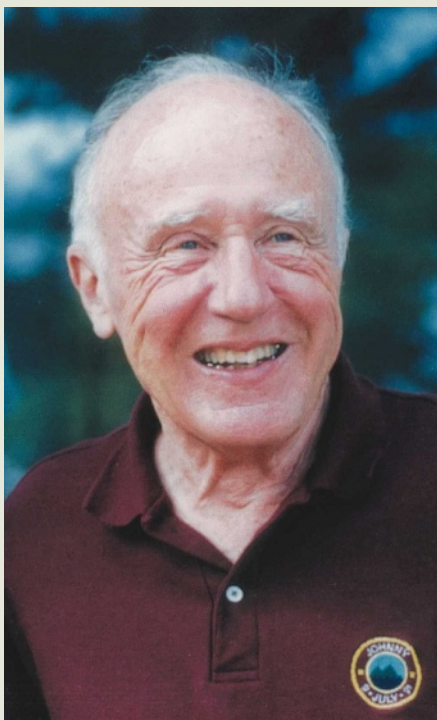
Theoretical physicist, inspired and inspiring teacher.

The fertile imagination of John Archibald Wheeler, who died on 13 April aged 96, roamed from the properties of atomic nuclei to the physics of nuclear explosions; from the quantum mechanics of exotic atoms to the philosophy of quantum measurements; and from the nature of black holes to the properties of space-time. The thread connecting these disparate areas was Wheeler's love of acquiring and imparting knowledge. Nowhere was that more evident than with the science of gravity, which Wheeler and his students — many of them themselves future leaders — took from a theoretical backwater to a vigorous area of scientific enquiry.

Wheeler loved 'what if' questions. What happens if electrons and their antiparticles, positrons, bubble up from the vacuum and cause light to be scattered by light? In 1934, just two years after the positron was discovered, Wheeler and Gregory Breit worked out the answer. Their work is still used to describe the scattering of energetic photons from distant galaxies. What if you replace the proton in a hydrogen atom by a positron, or the electron in an atom by a muon, its heavier cousin? Wheeler found that out in the 1940s. What if electrons and positrons are in fact the same particle, travelling in opposite directions in time? He put this question to his graduate student, Richard Feynman; it led to Feynman's powerful diagrammatic approach to quantum electrodynamics.

In 1939, nuclear fission was discovered. Wheeler, then at the start of a tenure of almost 40 years at Princeton University, collaborated with Niels Bohr to deliver an underlying theory. Soon after that, war broke out. Wheeler willingly went to work first on the American atomic bomb, and later the hydrogen bomb. He was a complex person: a patriot, he was offended by criticism of US military policy, for example in Vietnam. But he also encouraged one of us, Bill Unruh, to go to the University of California, Berkeley — he said that the many protests there at the time made it a stimulating place to work.

In 1952, Wheeler turned his attention to gravity. Could the malleable space-time of Einstein's general theory of relativity explain the characteristics of elementary particles? Wheeler's mantras, "charge without charge", "mass without mass" and "topology is too important to be left to the mathematicians", shocked and inspired others to draw out the physical implications of the theory. The rich harvest of physics and mathematics was displayed in the monumental treatise



*Gravitation*, written with two former students of his, Charles Misner and Kip Thorne. First published in 1973, it has not been out of print since.

One problem emphasized by the book was that of singularities. When a massive star exhausts its supply of nuclear fuel, it can collapse first to a 'black hole' — as Wheeler named it — and within that into a singularity where space and time cease to exist. As *Gravitation* was being published, another of Wheeler's students, Jacob Bekenstein, argued that black holes have thermodynamic properties such as energy and entropy. Three years later, Stephen Hawking showed that they have a temperature, emit radiation at that temperature, and so can evaporate. The strong observational evidence for black holes keeps Wheeler's fundamental questions alive — what is the meaning of a singularity; what is left after a black hole evaporates; what happens to information that is swallowed by a black hole.

Besides gravity, quantum mechanics was Wheeler's passion: "Why the quantum?" was another favourite question. He was especially fascinated by the idea that whether, looking at light, you see wave or particle behaviour depends on what experiment you perform. In a famous thought experiment, Wheeler considered what happened if the choice of experiment could be delayed until, assuming causality, it was too late for the decision to affect the way the light behaves.

The experiment has since been done, and it starkly displays the situation. In Wheeler's words, "We have a strange inversion of the normal order of time. We, now, by moving the mirror in or out, have a right to change what we say about the already past history of that photon." As a logical extension, Wheeler went on to ask whether our present experiments influence what we can say of the Universe in the distant past. Does the observer even in some way participate in the birth of the Universe?

Wheeler had no fear of such 'hard' questions. He infected others with his own intellectual courage, emboldening Hugh Everett, another of his doctoral students, to present his thesis positing the 'many worlds' interpretation of quantum mechanics, despite his, Wheeler's, own philosophical discomfort with it.

John Wheeler was an exuberant person. He happened to be in the office of one of us, Jim Peebles, when a student, Dan, came in with his new dissertation.

John: What's it about?

Dan: A search for correlations of position angles of galaxies.

John: Gödel would be interested.

Dan: Who's Gödel?

John: To say Gödel is the greatest logician since Aristotle would be to slight Gödel.

John picked up the phone, got Gödel on the line, and handed the phone to Dan. We don't know what Gödel made of this call (he had discovered a rotating-universe model), but John had got to do what he loved best: inspire a young person.

The rotation of galaxies. What Gödel's incompleteness theorem had to do with physics. The end of space-time. John Wheeler's eclectic passions are remote from the practical problems of society. But to him, fundamental scientific research was both a pleasure and a duty. His attitude was summed up in the dedication of *Gravitation*:

*To our fellow citizens  
Who, for love of truth,  
Take from their own wants  
By taxes and gifts,  
And now and then send forth  
One of themselves  
As a dedicated servant,  
To forward the search  
Into the mysteries and marvelous  
simplicities  
Of this strange and beautiful universe,  
Our home.*

**P. James E. Peebles and William G. Unruh**

P. James E. Peebles is in the Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. William G. Unruh is in the Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada. e-mails: pjep@princeton.edu; unruh@physics.ubc.ca

THE WHEELER FAMILY



# Rapid leukocyte migration by integrin-independent flowing and squeezing

Tim Lämmermann<sup>1</sup>, Bernhard L. Bader<sup>3</sup>, Susan J. Monkley<sup>4</sup>, Tim Worbs<sup>5</sup>, Roland Wedlich-Söldner<sup>2</sup>, Karin Hirsch<sup>1</sup>, Markus Keller<sup>3</sup>, Reinhold Förster<sup>5</sup>, David R. Critchley<sup>4</sup>, Reinhard Fässler<sup>1</sup> & Michael Sixt<sup>1</sup>

**All metazoan cells carry transmembrane receptors of the integrin family, which couple the contractile force of the actomyosin cytoskeleton to the extracellular environment. In agreement with this principle, rapidly migrating leukocytes use integrin-mediated adhesion when moving over two-dimensional surfaces. As migration on two-dimensional substrates naturally overemphasizes the role of adhesion, the contribution of integrins during three-dimensional movement of leukocytes within tissues has remained controversial. We studied the interplay between adhesive, contractile and protrusive forces during interstitial leukocyte chemotaxis *in vivo* and *in vitro*. We ablated all integrin heterodimers from murine leukocytes, and show here that functional integrins do not contribute to migration in three-dimensional environments. Instead, these cells migrate by the sole force of actin-network expansion, which promotes protrusive flowing of the leading edge. Myosin II-dependent contraction is only required on passage through narrow gaps, where a squeezing contraction of the trailing edge propels the rigid nucleus.**

The current model of metazoan cell migration is frequently described as a multistep cycle: F-actin polymerization at the cell front pushes out a membrane protrusion that subsequently becomes anchored to an extracellular substrate by transmembrane receptors of the integrin family. Integrins are dynamically coupled to the cytoskeleton and transduce the internal force that is generated when myosin II contracts the actin network. Contraction imposes retrograde pulling forces on the integrins, which in turn facilitates forward locomotion of the cell body<sup>1–3</sup>. The mammalian integrin family consists of 24 different functional heterodimers with individual binding specificities for cellular and extracellular ligands<sup>4</sup>. The integrin repertoire of each cell type defines which substrate it can use for this ‘haptokinetic’ (adhesion driven) mode of migration. Such intimate linkage between substrate-specific adhesion and migration restricts the migrating cells to preformed pathways, and thereby creates the determinism that is essential for many of the precise cell trafficking and positioning processes underlying compartmentalization and patterning during development and regeneration. As all mammalian cells with the exception of erythrocytes carry integrins on their surface, it is reasonable to assume that haptokinesis is a universal phenomenon<sup>4,5</sup>.

Leukocytes are outstanding cells, as they are scattered throughout the body and have the potential to infiltrate any type of tissue. Rather than following exact routes and being restricted to specific compartments, leukocytes frequently undergo stochastic swarming with single cell migration velocities that are up to 100 times faster than mesenchymal and epithelial cell types<sup>6,7</sup>. What makes leukocytes so quick and flexible? In contrast to slow cells, migrating leukocytes undergo frequent shape changes and were therefore morphologically described as ‘amoeboid’<sup>8</sup>. How these shape changes relate to actual locomotion is poorly investigated, and it is currently unknown if amoeboid migration represents a specialized strategy or just an accelerated variant of the above introduced migration paradigm that is only well established for slow cells. We therefore studied the inter-dependency of adhesion, protrusion and contraction in inflammatory cells. We demonstrate

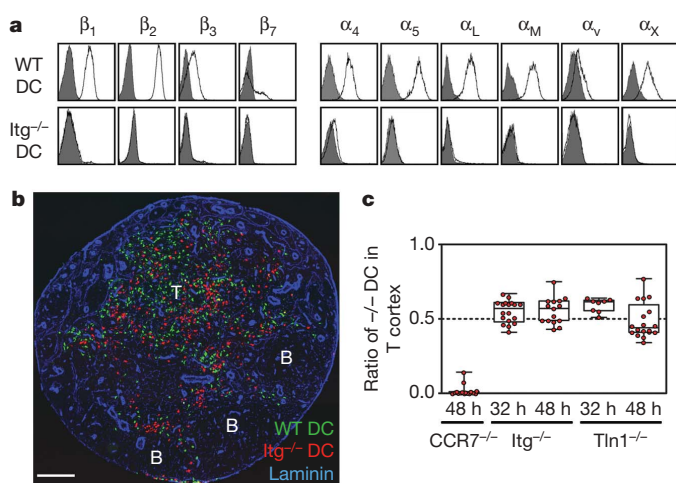
that integrin-mediated adhesion is only necessary to overcome tissue barriers like the endothelial layer, while interstitial migration is autonomous from the molecular composition of the extracellular environment. Such adhesion-independent migration is driven by protrusive flowing of the anterior actin network of the cell and supported by squeezing actomyosin contractions of the trailing edge to propel the rigid nucleus through narrow spaces.

## Dendritic cells migrate without integrins

As a model system to study interstitial leukocyte migration *in vivo*, we focused on dendritic cells (DCs), phagocytes that reside within peripheral tissues such as skin. They become activated upon wounding or infection, sample antigens and quickly migrate via the afferent lymphatic vessels into the draining lymph node where they act as antigen presenting cells<sup>9</sup>. DC migration is primarily guided by the two chemokines CCL19 and CCL21; these are expressed in lymphatic endothelium and the lymph nodes’ T-cell area, and bind the CC-chemokine receptor 7 (CCR7), which is upregulated on DCs upon activation<sup>10</sup>.

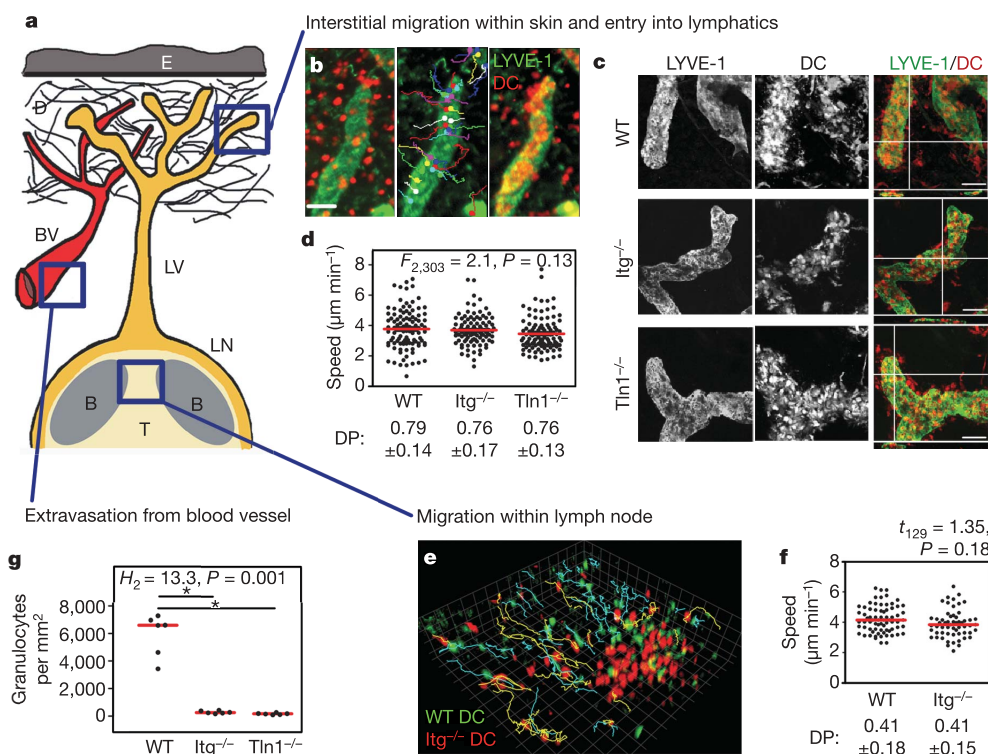
Flow cytometric analysis of DCs generated from mouse bone-marrow-derived stem cells revealed expression of the  $\beta_1$ ,  $\beta_2$ ,  $\beta_7$  and  $\alpha_v$  integrin families (Fig. 1a). To investigate the role integrins play in DC migration *in vivo*, we used a combinatorial mouse genetics approach to delete all integrin heterodimers from the surface of DCs, thereby generating phenotypically normal pan-integrin-deficient (integrin<sup>−/−</sup>) DCs (Fig. 1a, see Supplementary Material, Methods, and Supplementary Figs 1, 2, 4c). We co-injected 1:1 mixtures of differentially labelled integrin<sup>−/−</sup> and wild-type DCs into the dermis of mouse footpads and quantified their arrival in T-cell areas of the draining lymph nodes. Surprisingly, wild-type and integrin<sup>−/−</sup> DCs arrived and localized in the T-cell area in an indistinguishable manner (Fig. 1b, c, Supplementary Fig. 3), whereas CCR7<sup>−/−</sup> DCs, which cannot interpret the directional information, failed to enter the lymph node (Fig. 1c) as previously described<sup>10</sup>. We next targeted integrin functionality indirectly by deleting the talin1 gene in DCs

<sup>1</sup>Department of Molecular Medicine, <sup>2</sup>Junior Research Group Cellular Dynamics and Cell Patterning, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. <sup>3</sup>Department of Nutritional Medicine, Technische Universität München, Munich, 85350 Freising, Germany. <sup>4</sup>Department of Biochemistry, University of Leicester, Leicester LE1 9HN, UK. <sup>5</sup>Institute of Immunology, Hannover Medical School, 30625 Hannover, Germany.



**Figure 1 | Migration of integrin-deficient dendritic cells into lymph nodes.** **a**, Flow cytometric analysis of integrin subunits of wild-type (WT) and integrin-deficient ( $Itg^{-/-}$ ) dendritic cells (DCs) (integrin, unshaded; isotype control, grey-shaded). **b**, **c**, *In vivo* migration of DCs after subcutaneous injection into hind footpads of WT mice. At the time points indicated, arrival of differentially labelled DCs in the lymph node was analysed. **b**, Localization of DCs after 48 h. Composite of six separate images of a single lymph node. Scale bar, 200  $\mu$ m; B, B-cell follicle; T, T-cell cortex. **c**, Quantitative histology. Dotted line at 0.5, 50% of DCs that migrated into the lymph node are knockout DCs; Tln1, talin1. Red dots indicate single experiments (1 lymph node); box shows median, 25%, 75% percentile; whiskers show minimum, maximum.

(Supplementary Fig. 4a). The interaction of talin1 with integrin cytoplasmic domains is required for integrin activation, ligand binding and coupling of F-actin to adhesion sites<sup>11</sup>. Accordingly, talin1 gene ablation in DCs did not affect the levels of integrin expression, but completely abolished their ability to bind ligands (Supplementary Fig. 4b–d). Again, arrival of talin1<sup>-/-</sup> DCs in lymph nodes did not significantly differ from wild-type DCs (Fig. 1c).



**Figure 2 | Integrin-independent interstitial leukocyte migration *in vivo*.** **a**, Scheme of the dendritic cell (DC) migration path from the skin via lymphatic vessels to the lymph node; B, B-cell follicle; BV, blood vessel; D, dermis; E, epidermis; LN, lymph node; LV, lymphatic vessel; T, T-cell cortex.

**b–d**, Migration of DCs within the dermis of ear explants. **b**, Tracks of DCs (red) entering LYVE-1<sup>+</sup> LVs (green). **c**, Confocal images of whole mount ear explants 2 h after addition of DCs (red). DCs locate within the lumen of LVs; all scale bars, 50  $\mu$ m. **d**, Speed and directional persistence (DP, mean  $\pm$  s.d.) of single cells (dots) for wild-type (WT), integrin<sup>-/-</sup> ( $Itg^{-/-}$ ) and talin1<sup>-/-</sup> ( $Tln1^{-/-}$ ) DCs. Red line, mean. **e**, **f**, Intravital two-photon microscopy. **e**, 3D tracking of WT (green, blue line) and  $Itg^{-/-}$  (red, yellow line) DCs migrating in the interfollicular areas. **f**, Quantification of the average speed and DP (mean  $\pm$  s.d.) of tracked single cells (dots). Red line, mean. **g**, Quantification of granulocytes that extravasated from blood into ear dermis ( $n = 6$  for each mouse line). Red line, median. \* $P < 0.05$  (post hoc).

To allow for a detailed comparison of the cellular dynamics of wild-type and integrin<sup>-/-</sup> DCs within the different physiological environments of their migration route, we employed two *in situ* imaging approaches (Fig. 2). Representing the starting point of DC migration, the interstitial space of the dermis is dominated by fibrillar arrays of collagen bundles<sup>12</sup>. Using a newly developed set-up of *ex vivo* live cell imaging within explanted ear dermis, we could directly visualize DCs moving towards and entering the afferent lymphatic vessels by wide-field microscopy (Fig. 2b, c, Supplementary Fig. 5). Single cell tracking revealed no difference in the migratory behaviour of wild-type, integrin<sup>-/-</sup> and talin1<sup>-/-</sup> DCs (Fig. 2c, d).

The lymph node constitutes a fundamentally different environment for cellular migration, as it contains almost no freely accessible extracellular matrix molecules but is densely packed with lymphocytes and stroma cells<sup>13,14</sup>. To analyse the dynamics of intranodal DC migration from the subcapsular sinus towards the T-cell area, we employed a set-up of intravital two-photon microscopy of the popliteal lymph nodes<sup>15</sup>. Evaluating a range of motility parameters, including cell velocity, directional persistence, mean square displacement, motility coefficients and morphology, we found that the movement behaviour of wild-type and integrin<sup>-/-</sup> DCs was indistinguishable (Fig. 2e, f, Supplementary Videos 1–3, data not shown).

## 2D but not 3D migration is integrin-dependent

On their way into the lymph node, DCs do not cross significant tissue barriers, such as continuous endothelial or epithelial linings<sup>12</sup>. As these processes have been shown to be integrin-dependent<sup>16</sup>, we also tested for extravasation of integrin<sup>-/-</sup> leukocytes from the blood stream into the inflamed dermis, and found that it was indeed abolished (Fig. 2g, Supplementary Fig. 6). This finding corroborates the extravasation model in which integrin-mediated tight immobilization of leukocytes to the luminal endothelial surface is necessary to counteract the shear forces imposed by the blood flow<sup>16</sup>. In line with this *in vivo* finding, we found that integrin<sup>-/-</sup> DCs were entirely unable to adhere to and migrate on two-dimensional (2D) substrates. Gravity was obviously not sufficient to confine the non-adherent cells to planar surfaces and allow the transduction of traction forces



(Fig. 3a, b, Supplementary Fig. 4d, Supplementary Video 4). We conclude that extravasation requires integrin-mediated adhesion.

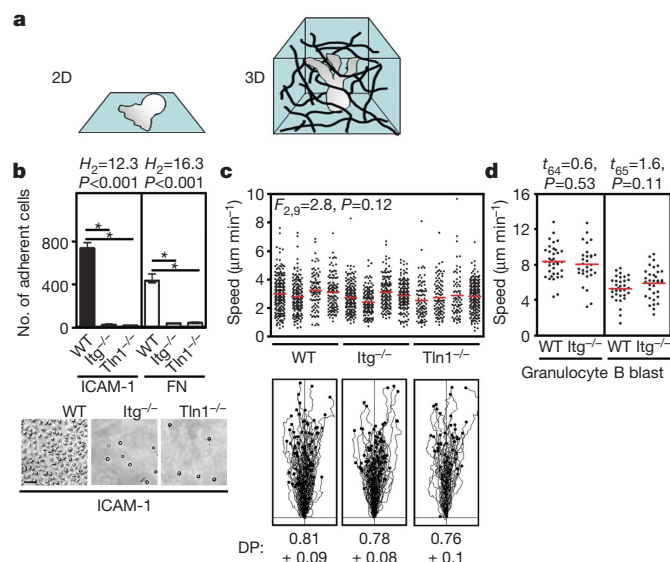
To better mimic the interstitial microenvironment, we established chemotaxis assays within artificial three-dimensional (3D) matrices of fibrin (a ligand for  $\beta_2$  and  $\beta_3$  integrins) and collagen I (a ligand for several members of the  $\beta_1$  integrin subfamily). In this system, DCs persistently migrated along soluble gradients of CCL19 and showed amoeboid morphology and velocities that were comparable to our (and previously published<sup>17</sup>) *in vivo* observations. In agreement with our *in vivo* data and in contrast to 2D migration, integrin<sup>-/-</sup> and talin1<sup>-/-</sup> DCs migrated with speed and directional persistence that were indistinguishable from wild-type cells (Fig. 3c, Supplementary Fig. 7a, Supplementary Video 5). We also observed integrin-independent migration for chemotaxing B cells and granulocytes, suggesting a broader prevalence of this functional principle (Fig. 3d, Supplementary Fig. 7b).

### Functional dissociation of front and back

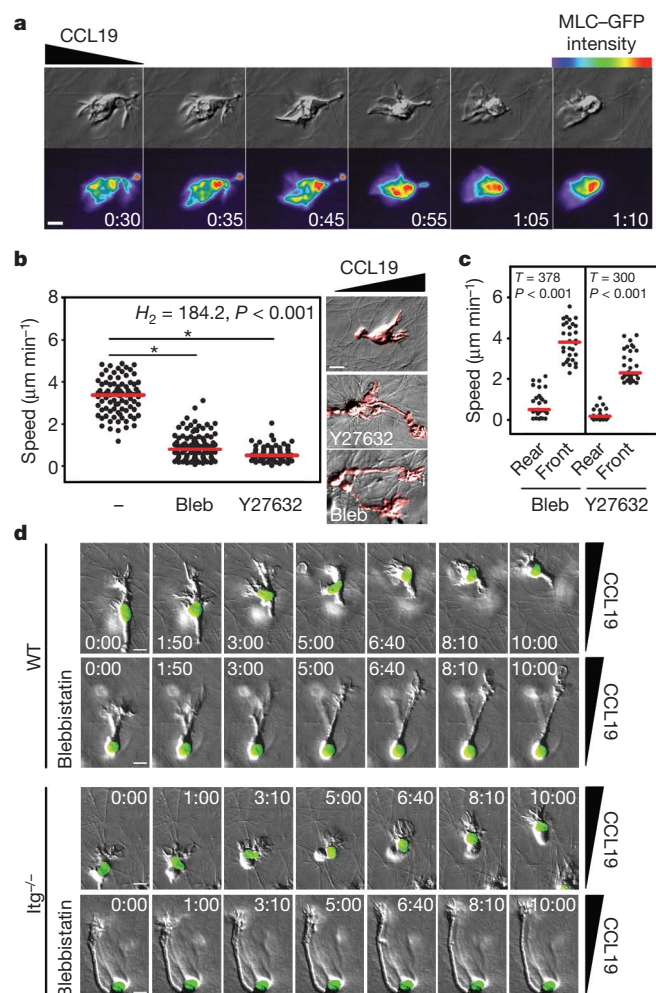
To mechanistically understand how cells can move in the absence of integrin-mediated traction forces, we compared the integrin-dependent movement pattern of fibroblasts with that of leukocytes and found a striking difference. Leading edge protrusions of fibroblasts translocated collagen fibres towards the cell body, indicating the presence of rearward pulling forces and demonstrating that protrusion, adhesion and contraction are tightly coupled in this cell type (Supplementary Video 6)<sup>18,19</sup>. In contrast, chemotaxing DCs protruded without signs of anterior pulling forces while the trailing edge displayed an irregularly alternating contractile pattern. Contraction was characterized by shrinkage of the cell rear with concomitant forward-streaming of cytoplasmic matter (Supplementary Video 7). When we dynamically visualized activated myosin II, the motor system mediating contraction<sup>1,2,20</sup>, by time-lapse imaging of DCs expressing a myosin light chain–GFP fusion protein, we found accumulation at the cell rear during contractile phases (Fig. 4a, Supplementary Fig. 8a, Supplementary Video 8). During non-contractile phases, the trailing edge remained motionless and

appeared to be passively dragged by the protruding cell front (Supplementary Video 7). These observations suggest that in contrast to slow moving cells, protrusion and contraction are spatio-temporally dissociated in leukocytes.

Receptor-mediated force transduction can only support forward movement if retrograde (contractile) but not anterograde (protrusive) forces are coupled to the environment<sup>2,3</sup>. To address how actomyosin contraction functionally contributes to leukocyte locomotion, we pharmacologically inhibited myosin II or its upstream-activator, Rho kinase. Irrespective of the presence of integrin function, either treatment severely reduced migration speeds of DCs, granulocytes and B cells (Fig. 4b, Supplementary Fig. 8e, data not shown). Nevertheless, the chemotactic gradient was still able to polarize the cell population. The leading edges of cells remained dynamic and protruded with normal speed towards the chemokine source. However, the cells were unable to move their trailing edges



**Figure 3 | Integrin-independent leukocyte migration in 3D networks *in vitro*.** **a**, 2D adhesion versus 3D migration of DCs. **b**, 2D adhesion on ICAM-1 and fibronectin (FN) 1 h after LPS stimulation. Top, quantification of the number of adherent cells (ICAM-1,  $n = 6$ ; FN,  $n = 8$  for each experiment). Bars, median values with interquartile range,  $*P < 0.05$  (post hoc). Bottom, morphology of wild-type (WT), integrin<sup>-/-</sup> (Itg<sup>-/-</sup>) and talin1<sup>-/-</sup> (Tln1<sup>-/-</sup>) DCs. **c**, **d**, Velocities of chemotaxing leukocytes in 3D collagen matrices. **c**, Top, DCs (4 experiments per group); **d**, granulocytes, B cells. Dots, single cells; red line, mean. **c**, Bottom, single cell tracks of chemotaxing DCs; values indicate mean DP  $\pm$  s.d.



**Figure 4 | Myosin II-dependent nuclear squeezing at the cell rear.** **a**, Sequence of a myosin light chain (MLC)–GFP expressing wild-type (WT) dendritic cell (DC) migrating towards CCL19 in a 3D collagen gel; upper row, differential interference contrast (DIC) microscopy; lower row, MLC–GFP intensity profiles (encoded in pseudo-colours with red representing highest levels). Time in min:s. **b**, Left, velocities of single DCs (dots; red line, median), and right, DC morphology (coloured red, DIC microscopy), chemotaxing towards CCL19 in 3D collagen gels upon pharmacological inhibition. **c**, Speed of cell bodies (rear) versus cell protrusions (front) of DCs in the presence of inhibitor; red line, median. Bleb, blebbistatin;  $*P < 0.05$  (post hoc). **d**, **e**, Time-lapse sequence of a WT (**d**) and integrin<sup>-/-</sup> (**e**) DC either with (lower row) or without (upper row) blebbistatin; DIC microscopy, nuclei (green). Time in min:s; scale bars, 5 μm (**a**), 10 μm (**b**, **d**, **e**).

(Fig. 4b–e, Supplementary Fig. 8b–d, Supplementary Video 9). This functional dissociation between front and back caused up to 30-fold cell elongation, and demonstrates that the leading edge migrates autonomously and without a need for receptor-mediated coupling of contractile forces to the extracellular matrix.

### Trailing edge contraction propels the nucleus

How does trailing edge contraction contribute to cell body locomotion? In fibroblasts and leukocytes moving on 2D substrates, actomyosin contraction at the back is required to disassemble receptor binding-sites and subsequently retract the membrane<sup>2,21,22</sup>. Hence, blocking contraction causes membrane tethers at the trailing edge. Consistent with an adhesion-independent migration mode, we could not observe membrane tethers in myosin II-inhibited leukocytes migrating in 3D gels. We considered that the elongated phenotype with its rounded back was caused by the inability to move an

internal resistance through narrow gaps within the gel. As the nucleus is the least elastic cellular compartment<sup>23</sup>, we visualized DNA within chemotaxing DCs and granulocytes. In untreated cells, we observed continuous shape changes of nuclei, indicating deformation forces, while upon myosin II inhibition, spherical nuclei were immobilized at the rear ends of the cells (Figs 4d, e, Supplementary Fig. 8b–d, Supplementary Videos 10, 11).

### Protrusion drives basal locomotion

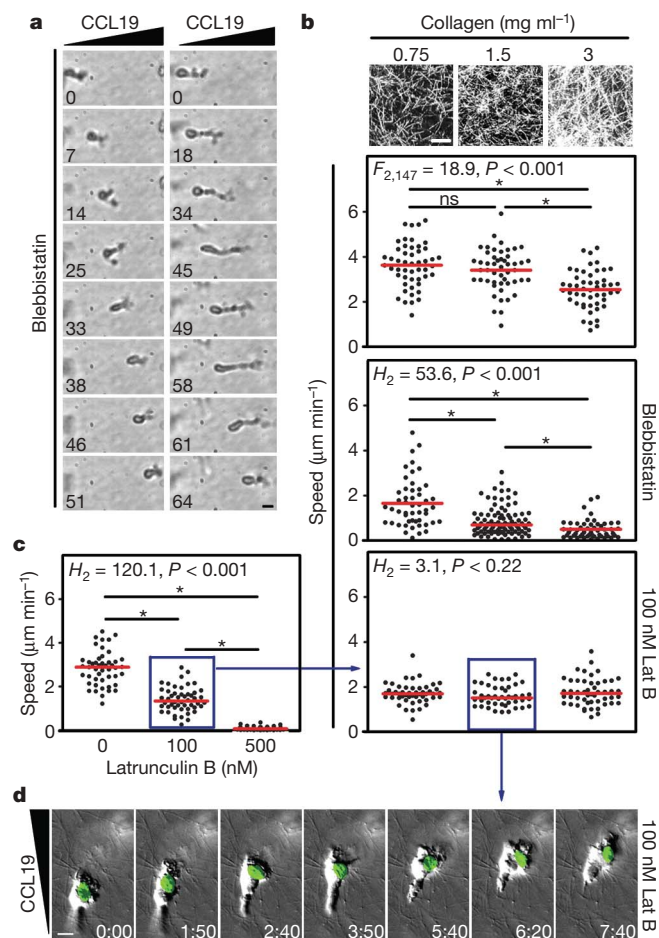
Although ‘locked’ nuclei caused migration arrest in most cells, a few still showed locomotion. This residual migration was often characterized by cell body elongation and dragging of the nucleus (Fig. 5a, Supplementary Video 12). Such non-contractile movement was in line with our previous observations showing motile phases in the absence of trailing edge contractions. We hypothesized that such myosin II-independent migration might occur in areas of the collagen gel where external resistance was low due to increased spacing of the collagen fibres. To establish the relationship between internal contractile force and external resistance, we analysed DC chemotaxis in collagen gels of varying fibre spacing (Fig. 5b). In all collagen densities, myosin II-inhibited DCs were slower than untreated cells but importantly, they ‘caught up’ at lowest gel densities (Fig. 5b, Supplementary Video 13). A comparison of instantaneous velocities further showed that myosin II inhibition did not prevent DCs from reaching the same peak values as untreated cells (Supplementary Fig. 9a–c).

We conclude that, in contrast to the ‘blebbing’ model of amoeboid cell migration<sup>24</sup>, cortical actin network contraction does not mediate locomotion itself but rather facilitates a protrusive mode of migration in confined environments where protrusion alone is unable to act against counter-forces. So in constricted areas, the cell overcomes internal and external resistance (rigidity of the nucleus and fibre density) by contraction (myosin II). Contractile force deforms the resistance and facilitates a purely protrusive mode of migration.

To challenge the concept of protrusive movement, we treated chemotaxing DCs with latrunculin B, which interferes with actin polymerization by depleting the pool of available functional G-actin monomers<sup>25</sup>. We found that this treatment reduced cell migration velocity in a dose-dependent manner (Fig. 5c). At intermediate concentrations, the trailing edge remained contractile and the nucleus was ‘pushed’ to the cell front (Fig. 5d, Supplementary Video 14). Consistent with a model where protrusion determines the speed of cell migration while contraction is only activated to overcome external resistances, the speed reduction of latrunculin B-treated cells was independent of gel density (Fig. 5b, c).

### Discussion

We show that leukocytes migrate in the absence of specific adhesive interactions with the extracellular environment. This subversion of the metazoan principle makes them autonomous from the tissue context, and allows them to quickly and flexibly navigate through any organ without adaptations to alternating extracellular ligands. This strategy is in stark contrast to the haptokinetic migration principle that leads along preformed pathways and therefore promotes deterministic positioning. Adhesion independency better suits stochastic movement or chemotaxis where cells randomly migrate or follow soluble cues. Astonishingly, the protrusive flow of F-actin appears sufficient to drive rapid leukocyte locomotion in environments with large pore-size. This resembles the locomotion principle of nematode sperm cells that is entirely driven by treadmilling polymers of major sperm protein<sup>26</sup>. Only in narrow areas do leukocytes activate the contractile module to squeeze and propel the internal resistance of the nucleus in a manner resembling neuronal nucleokinesis<sup>27</sup>. This ‘flowing and squeezing’ migration model fulfils a central requirement for immune cell movement: the pericellular environment is transiently deformed but never digested<sup>28</sup> or



**Figure 5 | Myosin II-independent protrusive migration of dendritic cells.** **a**, Two types of residual migration of wild-type (WT) dendritic cells (DCs) treated with blebbistatin in a standard 3D collagen gel. DCs either drag the nucleus behind with elongated appearance and low speed (right column) or appear morphologically normal with high speed (left column). Time in min. **b**, Top panel, confocal reflection microscopy of 3D collagen networks of different densities. Lower panels, velocities of single chemotaxing WT DCs (dots) in collagen gels with varying densities. Upper graph, no inhibitor; red line, mean. Middle graph, blebbistatin; red line, median. Bottom graph, 100 nM latrunculin B (partial actin depolymerization); red line, median. **c**, Velocity of single WT DCs (dots) in the presence of different concentrations of latrunculin B (1.6 mg ml<sup>-1</sup> collagen); red line, median. In **b** and **c**, \*P < 0.05 (post hoc). **d**, Time-lapse sequence of a 100 nM latrunculin B-treated WT DC; DIC microscopy, nuclei (green). Time, min:s. Blue boxes (in **b** and **c**) highlight experimental conditions that were used for follow-up experiments indicated by the blue arrow. Scale bars: 10 μm (**a**, **d**), 50 μm (**b**).



otherwise permanently remodelled which avoids 'collateral damage', caused by infiltrating cells.

The fact that leukocytes are able to move autonomously generates a new level of regulatory possibilities. Because surface bound chemokines and other immobilized extracellular signals do trigger integrin affinity<sup>16</sup> (unlike soluble chemokines), leukocyte integrins should no longer be viewed as force transducers during locomotion but as switchable immobilizing anchors that stop, slow down or confine high intrinsic motility to specifically assigned surfaces<sup>29,30</sup>. The role of integrins is therefore mostly to mediate retention, invasion, cell-cell communication and cell-cell adhesion<sup>31</sup>.

## METHODS SUMMARY

**Generation of integrin-deficient leukocytes.** Integrins and talin1 were targeted by generating mice with the genotype  $\alpha_v^{\text{flox/flox}}$  (Supplementary Fig. 1),  $\beta_1^{\text{flox/flox}}$  (ref. 32),  $\beta_2^{-/-}$  (ref. 33),  $\beta_7^{-/-}$  (ref. 34),  $\text{Mx1Cre}^{+/-}$  (ref. 35) and talin1 (ref. 36),  $\text{Mx1Cre}^{+/-}$ , respectively. Cre expression in the haematopoietic system was induced by intraperitoneal injection of 250  $\mu\text{g}$  Poly (I)·Poly (C) (Amersham Biosciences). 10–14 d after knock-out induction DCs were generated from bone marrow suspension and matured with lipopolysaccharide (LPS). The DC culture was depleted for granulocytes and remaining integrin-positive contaminants by magnetic sorting (Miltenyi Biotec). DCs used for migration assays were >99% enriched for  $\beta_1$  and  $\alpha_v$  integrin knockout cells.

**In situ live cell imaging.** For dermal *ex vivo* microscopy, mouse ears were mechanically separated in dorsal and ventral halves, fluorescently stained with LYVE-1 antibody and immobilized with the epidermal side down. The dermis was overlaid with fluorescently labelled DCs and time lapse movies were recorded with an automated Leica MZ 16 FA stereomicroscope (Visitron Systems).

**In vitro 3D chemotaxis assays.** Cells were suspended in PureCol (INAMED) and cast in custom built migration chambers (standard collagen concentration: 1.6  $\text{mg ml}^{-1}$ ). After polymerization, gels were overlaid with culture medium containing the chemoattractant (CCL19, C5a and CXCL13 for DCs, granulocytes and B cells, respectively) and subsequently imaged using wide-field fluorescence video microscopes with differential interference contrast (Zeiss). For nucleus visualization, SYTO-dyes (Invitrogen) were used. For visualization of myosin light chain, DCs were nucleofected using Amaxa technology. Inhibitors were titrated and used 50  $\mu\text{m}$  blebbistatin (Sigma) and 30  $\mu\text{m}$  Y27632 (Calbiochem).

**Statistical analysis.** *t*-tests and analysis of variance (ANOVA) were performed after data were confirmed to fulfil the criteria of normal distribution and equal variance, otherwise Kruskal–Wallis tests or Mann–Whitney U-tests were applied. If overall ANOVA or Kruskal–Wallis tests were significant, we performed a post hoc test. Analyses were performed with Sigma Stat 2.03. For further statistical details, see Supplementary Table.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 November 2007; accepted 6 March 2008.

- Giannone, G. *et al.* Lamellipodial actin mechanically links myosin activity with adhesion-site formation. *Cell* **128**, 561–575 (2007).
- Lauffenburger, D. A. & Horwitz, A. F. Cell migration: A physically integrated molecular process. *Cell* **84**, 359–369 (1996).
- Mitchison, T. J. & Cramer, L. P. Actin-based cell motility and cell locomotion. *Cell* **84**, 371–379 (1996).
- Hynes, R. O. Integrins: Bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).
- Hynes, R. O. & Zhao, Q. The evolution of cell adhesion. *J. Cell Biol.* **150**, 89–96 (2000).
- Friedl, P. Preshaping and plasticity: Shifting mechanisms of cell migration. *Curr. Opin. Cell Biol.* **16**, 14–23 (2004).
- Pittet, M. J. & Mempel, T. R. Regulation of T-cell migration and effector functions: Insights from *in vivo* imaging studies. *Immunol. Rev.* **221**, 107–129 (2008).
- de Bruyn, P. P. H. The amoeboid movement of the mammalian leukocyte in tissue culture. *Anat. Rec.* **95**, 117–192 (1946).
- Reis e Sousa, C. Dendritic cells in a mature age. *Nature Rev. Immunol.* **6**, 476–483 (2006).
- Förster, R. *et al.* CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23–33 (1999).
- Calderwood, D. A. & Ginsberg, M. H. Talin forges the links between integrins and actin. *Nature Cell Biol.* **5**, 694–697 (2003).
- Stoitzner, P., Pfaller, K., Stössel, H. & Romani, N. A close-up view of migrating Langerhans cells in the skin. *J. Invest. Dermatol.* **118**, 117–125 (2002).
- Gretz, J. E., Anderson, A. O. & Shaw, S. Cords, channels, corridors and conduits: Critical architectural elements facilitating cell interactions in the lymph node cortex. *Immunol. Rev.* **156**, 11–24 (1997).
- Lämmermann, T. & Sixt, M. The microanatomy of T-cell responses. *Immunol. Rev.* **221**, 26–43 (2008).
- Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154–159 (2004).
- Alon, R. & Dustin, M. L. Force as a facilitator of integrin conformational changes during leukocyte arrest on blood vessels and antigen-presenting cells. *Immunity* **26**, 17–27 (2007).
- Lindquist, R. L. *et al.* Visualizing dendritic cell networks *in vivo*. *Nature Immunol.* **5**, 1243–1250 (2004).
- Meshel, A. S., Wei, Q., Adelstein, R. S. & Sheetz, M. P. Basic mechanism of three-dimensional collagen fibre transport by fibroblasts. *Nature Cell Biol.* **7**, 157–164 (2005).
- Vanni, S., Lagerholm, B. C., Otey, C., Taylor, D. L. & Lanni, F. Internet-based image analysis quantifies contractile behavior of individual fibroblasts inside model tissue. *Biophys. J.* **84**, 2715–2727 (2003).
- Medeiros, N. A., Burnette, D. T. & Forscher, P. Myosin II functions in actin-bundle turnover in neuronal growth cones. *Nature Cell Biol.* **8**, 215–226 (2006).
- Hogg, N., Laschinger, M., Giles, K. & McDowall, A. T-cell integrins: More than just sticking points. *J. Cell Sci.* **116**, 4695–4705 (2003).
- Morin, N. A. *et al.* Nonmuscle myosin heavy chain IIA mediates integrin LFA-1 de-adhesion during T lymphocyte migration. *J. Exp. Med.* **205**, 195–205 (2008).
- Hu, S., Chen, J., Butler, J. P. & Wang, N. Prestress mediates force propagation into the nucleus. *Biochem. Biophys. Res. Commun.* **329**, 423–428 (2005).
- Paluch, E., Sykes, C., Prost, J. & Bornens, M. Dynamic modes of the cortical actomyosin gel during cell locomotion and division. *Trends Cell Biol.* **16**, 5–10 (2006).
- Morton, W. M., Ayscough, K. R. & McLaughlin, P. J. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. *Nature Cell Biol.* **2**, 376–378 (2000).
- Bottino, D., Mogilner, A., Roberts, T., Stewart, M. & Oster, G. How nematode sperm crawl. *J. Cell Sci.* **115**, 367–384 (2002).
- Schaar, B. T. & McConnell, S. K. Cytoskeletal coordination during neuronal migration. *Proc. Natl Acad. Sci. USA* **102**, 13652–13657 (2005).
- Wolf, K., Müller, R., Borgmann, S., Bröcker, E. B. & Friedl, P. Amoeboid shape change and contact guidance: T-lymphocyte crawling through fibrillar collagen is independent of matrix remodeling by MMPs and other proteases. *Blood* **102**, 3262–3269 (2003).
- Auffray, C. *et al.* Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior. *Science* **317**, 666–670 (2007).
- Phillipson, M. *et al.* Intraluminal crawling of neutrophils to emigration sites: A molecularly distinct process from adhesion in the recruitment cascade. *J. Exp. Med.* **203**, 2569–2575 (2006).
- Sixt, M., Bauer, M., Lämmermann, T. & Fässler, R. Beta1 integrins: Zip codes and signaling relay for blood cells. *Curr. Opin. Cell Biol.* **18**, 482–490 (2006).
- Potocnik, A. J., Brakebusch, C. & Fässler, R. Fetal and adult hematopoietic stem cells require beta1 integrin function for colonizing fetal liver, spleen, and bone marrow. *Immunity* **12**, 653–663 (2000).
- Wilson, R. W. *et al.* Gene targeting yields a CD18-mutant mouse for study of inflammation. *J. Immunol.* **151**, 1571–1578 (1993).
- Wagner, N. *et al.* Critical role for beta7 integrins in formation of the gut-associated lymphoid tissue. *Nature* **382**, 366–370 (1996).
- Kühn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Petrich, B. G. *et al.* Talin is required for integrin-mediated platelet function in hemostasis and thrombosis. *J. Exp. Med.* **204**, 3103–3111 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Cremer for help with statistical analysis, Z. Werb and P. Friedl for critical reading of the manuscript, and M. Bauer for technical support. This work was financed by the German Research Foundation (DFG), the Austrian Science Foundation (FWF) and the Max Planck Society. Work in D.R.C.'s laboratory was supported by the Wellcome Trust.

**Author Contributions** T.L. and M.S. designed and performed the experiments and analysed the data. M.S. wrote the paper. T.W. and R.F. performed intravital microscopy in lymph nodes. B.L.B. and M.K. generated the integrin  $\alpha_v$  mouse. S.J.M. and D.R.C. generated the talin1 mouse. R.F. generated the integrin  $\beta_1$  and the quadruple integrin knockout mouse and provided general support. K.H. assisted with experiments. R.W.S. contributed to data analysis and experimental design.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.S. (sixt@biochem.mpg.de).

## METHODS

**Generation of integrin  $\alpha_v$ <sup>+/-lox</sup> mice.** To generate a conditional floxed  $\alpha_v$  allele, a targeting vector spanning 8.2 kb of the murine  $\alpha_v$  integrin locus<sup>37</sup> was constructed. First, the 2.2 kb genomic subclone pXB2.2 encoding the 5' upstream region, exon 1 and part of the intron 1 of the  $\alpha_v$  gene was used to introduce two loxP-sites. The 5'-loxP1-site including an EcoRI-restriction site to facilitate the identification of the targeted locus was cloned into the HindIII-site in the 5'-upstream region (nucleotide position, nt-pos., -310) of the  $\alpha_v$  gene. The 3'-loxP2-site was introduced into the BamHI-site in intron 1 (nt-pos. +528). The genomic subclone pBS6 harbouring this BamHI-site at its 5'-end and ~6 kb 3'-downstream intron 1 sequences was used to clone a *pgk-neo;pgk-tk*-tandem gene selection-cassette (*neoTK*) flanked by two *frt*-sites (provided by G. R. Martin) into the BamHI-site. These two modified genomic regions of the  $\alpha_v$  gene were then fused at the corresponding BamHI-site to obtain the  $\alpha_v$  gene targeting vector. ES cell transfection, selection and screening was carried out as described<sup>37</sup> except that R1 ES cells<sup>38</sup> were used. Homologously targeted clones were identified by Southern blotting of EcoRI- or BamHI-digested DNA with 5'- or 3'-external probes A or B, respectively. To excise the neoTK selection cassette, targeted ES cells (*flneoTK*) were transiently transfected with the CAGGS-FLPe-expression construct<sup>39</sup> (provided by A. F. Stewart) and counterselected with ganciclovir. Loss of the selection cassette in the  $\alpha_v$  floxed allele was then confirmed by Southern blotting. ES cell clones were injected into blastocysts of C57BL/6 mice, and resulting germline transmitting male chimaeras were mated to C57BL/6 females to generate heterozygous floxed  $\alpha_v$  mice and wild-type F1 mice following protocols as described<sup>37</sup>. Two independent lines carrying the  $\alpha_v$ <sup>+/-lox</sup> allele were established and initially maintained on a 129SV/C57BL/6 mixed background.  $\alpha_v$ <sup>+/-lox</sup> and  $\alpha_v$ <sup>fl/fl</sup> mice show no obvious phenotype. Southern blot and PCR analyses: wild-type and mutant  $\alpha_v$  alleles were assessed by Southern blot hybridization and/or PCR from DNA isolated from ES cells, mouse tail biopsies or from yolk sacs of embryos as described<sup>37</sup>. In brief, 10 µg of genomic DNA was digested with EcoRI or BamHI and analysed using radio-labelled 5'-external probe A (419 bp-EcoRI/XbaI DNA-fragment) or a 3'-external probe B (690 bp-SphI/EcoRI DNA-fragment), respectively (see Supplementary Fig. 1). Genomic DNA digested with EcoRI was used to confirm the presence or absence of the selection cassette with probe C encoding a 800 bp-PstI fragment of the neomycin-resistance gene. For genotyping wild-type (WT, +) and floxed  $\alpha_v$  allele (floxed) as well as the *cre*-mediated knockout allele ( $\alpha_v$ <sup>Δex1</sup>), PCR amplifications with genomic DNA and the specific primer set (see below) were performed. The positions of the primers are located in close vicinity to the loxP-sites. PCR amplifications using *Taq*-polymerase and standard reaction buffer adjusted to a final concentration of 1 mM MgCl<sub>2</sub> were carried out as follows. PCR samples were initially denatured at 94 °C for 3 min, followed by 35 cycles of amplification (94 °C for 1 min, 60 °C for 2 min, 72 °C for 3 min), and a final 10 min extension at 72 °C followed. PCR amplification products were separated through agarose gel electrophoresis. The allele-specific primer sequences, pairs and orientation (forward, Fw; reverse, Rev), the primer pairs and their corresponding PCR-product sizes are as follows: Primer-set (F30: 5'-AGGATGGAAGGGGAGGGAATG-3' (Fw; WT), F20: 5'-CTTGACCGCAA-GCGCACAGCACAG-3' (Fw; WT), I2: 5'-CTGGATGCTGAGTGTCAGGT-3' (Rev; WT)), PCR products (F20/I2: 499 bp (wild-type  $\alpha_v$  allele, WT); F20/I2: 614 bp (floxed  $\alpha_v$  allele,  $\alpha_v$ <sup>fl/fl</sup>); F30/I2: 339 bp (*cre*-mediated  $\alpha_v$  knockout allele,  $\alpha_v$ <sup>Δex1</sup>)).

**Mouse strains.** Mice with constitutive deletions of integrin  $\beta_7$ , integrin  $\beta_2$ , CCR7 and conditional  $\beta_1$ <sup>fl/fl</sup> mice were described<sup>10,32-34</sup>. All mice were kept on a mixed 129SV/C57BL/6 genetic background. The integrin targeted strains were inter-crossed with Mx1Cre transgenic mice to generate  $\alpha_v$ <sup>fl/fl</sup>,  $\beta_1$ <sup>fl/fl</sup>,  $\beta_2$ <sup>-/-</sup>,  $\beta_7$ <sup>-/-</sup>, Mx1Cre<sup>+/-</sup> animals. *Talin1*<sup>fl/fl</sup> mice<sup>36</sup> were crossed with Mx1Cre transgenic mice to obtain *talin1*<sup>fl/fl</sup>, Mx1Cre<sup>+/-</sup> mice. All control animals were on a mixed 129SV/C57BL/6 genetic background. The mice were bred in a conventional animal facility at the Max Planck Institute of Biochemistry, and according to the local regulations.

**Induction of the haematopoietic knockout.** At an age between 8 and 12 weeks, mice of the desired genotype received a single intraperitoneal injection of 250 µg Poly (I)·Poly (C) (Amersham Biosciences), diluted in 0.5 ml phosphate buffered saline (PBS). Mice were killed 10–14 d after injection and knockout efficiency of haematopoietic precursors was estimated by flow cytometric analysis of peripheral blood platelets<sup>40</sup>. Bone marrow (BM) of mice with >85% integrin  $\beta_1$ -negative platelets was processed for dendritic cell (DC) culture or cell isolation. Platelet knockout efficiencies usually corresponded with knockout efficiencies of DCs and granulocytes, whereas peripheral B cells frequently showed lower efficiency.

**Cell generation, separation and purification.** DCs were generated from flushed BM suspension as described previously<sup>41</sup>. At day 8–10 of culture, 200 ng ml<sup>-1</sup>

LPS (Sigma-Aldrich; *E. coli* 0127:B8) was added overnight and the pan-integrin-deleted DCs were subsequently depleted of remaining Gr-1<sup>+</sup> and integrin  $\beta_1$ <sup>+</sup> contaminants. For depletion, cells were incubated with biotinylated antibodies against Gr-1 (RB6-8C5) and integrin  $\beta_1$  (Ha2/5) (both BD Pharmingen), followed by streptavidin-microbead separation according to the manufacturers protocol (Miltenyi Biotech). The negatively selected DCs used for migration assays were >99% enriched for  $\beta_1$  integrin knockout cells and  $\beta_1$ -deficient cells were always 100% deleted for integrin  $\alpha_v$ . After enrichment the cells are referred to as integrin<sup>-/-</sup> DCs. *Talin1*<sup>-/-</sup> DCs were generated identically and depleted of Gr-1<sup>+</sup> contaminants. To determine knockout efficiencies, DC lysates were loaded on 12–15% gradient gels for electrophoresis under reduced conditions and western blotting. Talin was detected with mouse anti-pan-talin antibody (T3287, Sigma), and the loading control was performed with rabbit anti-mouse actin antibody (A2066, Sigma). Integrin  $\beta_1$  was detected with the rabbit polyclonal anti-integrin  $\beta_1$  antibody 1244<sup>42</sup>. Integrin  $\alpha_v$  was detected with a rabbit polyclonal anti-integrin  $\alpha_v$  antibody (Chemicon). Bound primary antibodies were detected with anti-mouse or anti-rabbit horse radish peroxidase (HRP) or anti-rabbit HRP (both Bio-Rad) as secondary antibodies, respectively. DCs with no detectable talin stain were chosen for experiments, guaranteeing that the cells were deleted for talin1 and that talin2 was neither expressed nor induced as a compensation mechanism in DCs. Granulocytes were isolated by staining whole bone marrow suspension with anti-Gr-1 PE or anti-Gr-1 PE/anti- $\beta_1$  FITC and subsequent fluorescent activated cell sorting using a FACSaria (BD Biosciences), revealing >96% purity. Before incorporation into the gels, granulocytes were treated for 30 min with 20 ng ml<sup>-1</sup> TNF- $\alpha$  (Roche) at 37 °C. B lymphoblasts were generated by mincing spleens through a 70 µm nylon cell strainer (BD Pharmingen), lysing erythrocytes with ACK lysis buffer (0.15 M NH<sub>4</sub>Cl, 1 mM KHCO<sub>3</sub>, and 0.1 mM EDTA), and subsequent magnetic cell sorting (MACS) with mouse CD43 MicroBeads on LS columns (Miltenyi). The negatively sorted B cells were stimulated with 10 µg ml<sup>-1</sup> LPS (Sigma) for 72 h at 37 °C, 5% CO<sub>2</sub>. Streptavidin-microbead separation of integrin  $\beta_1$ <sup>+</sup> cells (see DCs) revealed >98% pan-integrin-deficient B lymphoblasts that we refer to as integrin<sup>-/-</sup> B lymphoblasts.

**Flow cytometry.** Mature DCs were identified as CD11c<sup>+</sup>, MHCII<sup>high</sup> cells for wild-type and talin1-deficient cells. For integrin<sup>-/-</sup> cells, MHCII<sup>high</sup> cells corresponded to mature DCs as they carried enhanced levels of maturation markers CD86 and CD40. B lymphoblasts were identified as B220<sup>+</sup> cells, granulocytes as Gr-1<sup>+</sup> cells. Employed antibodies were against the mouse antigens: Gr-1 biotin and PE (RB6-8C5), B220 FITC (RA3-6B2), CD11c FITC and biotin (HL3), I-A/I-E biotin (2G9) and I-A/I-E PE (M5/114.15.2) (all from BD Pharmingen). For integrin expression profiles, the above cell markers were combined with antibodies raised against the mouse integrin chains:  $\alpha_2$  FITC (Ha 1/29),  $\alpha_{IIb}$  FITC (MWReg30),  $\alpha_M$  PE (M1/70),  $\alpha_4$  PE (9C10),  $\alpha_5$  PE (5H10-27),  $\alpha_6$  PE (GoH3),  $\alpha_E$  PE (M290),  $\alpha_v$  (RMV-7),  $\beta_1$  biotin (Ha2/5),  $\beta_2$  biotin (C71/16),  $\beta_3$  biotin (2C9.G2),  $\beta_7$  biotin (2C3.G2) (all from Pharmingen),  $\alpha_1$  FITC (HMApha1),  $\beta_4$  PE (346-11A) (both from Serotec),  $\alpha_L$  biotin (M17/4) (eBiosciences) and  $\beta_1$  PE (HMBeta1-1) (BioLegend). The following isotype-matched antibody controls were employed: rat IgG2a PE (B39-4, R35-95), rat IgG1 PE (R3-34), hamster IgG2 FITC (Ha4/8), hamster IgM FITC (G235-1) (all Pharmingen), rat IgG2a biotin, rat IgG2b FITC (eB149/10H5), rat IgG1 FITC (all eBiosciences), hamster IgG PE (HTK888), hamster IgM biotin (HTK204) (both BioLegend) and rat IgG2a PE (YTH7.1.3) (Serotec). Biotinylated antibodies were detected with secondary streptavidin-Cy5 (Jackson). Flow cytometric analysis was performed with a FACScalibur and CellQuest Pro Software (BD Biosciences).

**Quantitative RT-PCR.** RNA was isolated from sorted DCs (see above) using the Absolutely RNA Microprep Kit (Stratagene) and the RNeasy Mini Kit (Qiagen). One microlitre cDNA, generated from 15 ng RNA using the iScript cDNA Synthesis Kit (Bio-Rad), was subjected to real-time PCR with the iQ SYBR Green Supermix (Bio-Rad) on the iCycler (Bio-Rad). Primers for integrin  $\beta_1$  5'-AGACTTCGCGATTGGCTTTG-3' and 5'-GCTGGTGACAGTTTGTTCAC-3', and integrin  $\alpha_v$  5'-CAAGCTCACTCCCATCAC-3' and 5'-GGGTGTCTT-GATTCTCAAAGGG-3' were used. Integrin gene expression in wild-type and integrin<sup>-/-</sup> DCs was quantified with the Gene Expression Analysis Program for the iCycler iQ Real-Time PCR Detection System (Bio-Rad). GAPDH was used to normalize gene expression, and integrin expression levels in wild-type DCs were set to 1.

**In vivo migration assay.** Gr-1-depleted wild-type and integrin<sup>-/-</sup> DCs were labelled with 3.5 µM cell-permeable 5-(and-6)-carboxyfluorescein succinimidyl ester (CFSE) or 1 µM tetramethylrhodamine (TAMRA) (both Invitrogen), respectively, and vice versa (the fluorescent labels did not influence DC migration (not shown)). 1 × 10<sup>6</sup> DCs at a 1:1 ratio were suspended in 20 µl PBS and injected subcutaneously into the hind footpads of C57BL/6 mice. 32 and 48 h after injection, mice were killed by CO<sub>2</sub> suffocation, popliteal lymph nodes frozen in cryomatrix (Thermo) and sectioned in 12 µm slices. To visualize the



lymph node compartments, methanol-fixed cryosections were stained after blocking with 1% BSA in PBS with pan-laminin antibody (L9393, Sigma) and secondary anti-rabbit Cy5 (Jackson). Three different layers of the T cell cortex of each lymph node were documented using an Axio Imager upright microscope (Zeiss). Quantification of DCs within the T cell cortex was performed with MetaMorph (Molecular Devices). Using inverse thresholding, the greyscale images of both fluorescence channels were separately binarized. DCs appeared as dots that were quantified using the morphometric analysis tool. The ratio of migrated knockout DCs was calculated as the number of knockout cells divided by the total number of DCs in the T cell cortex, with 0.5 as the value where 50% of knockout cells and 50% wild-type cells located in the centre of the lymph node. To confirm that fluorescent cells in the lymph node represent DCs that migrated from the injection site (Supplementary Fig. 3), DCs were labelled with 3  $\mu\text{M}$  Cell Trace Oregon Green 488 (Invitrogen).  $1 \times 10^6$  DCs of either wild-type or integrin<sup>-/-</sup> DCs were injected as described above and 48 h later DCs were identified as green cells expressing CD86 by flow cytometric analysis.

**Ex vivo wide-field microscopy in ear dermis.** Ears of killed C57BL/6 mice were removed and mechanically split into dorsal and ventral halves. During separation the dermis detached from the cartilage layer that remained on one of the halves. The cartilage free half was left unfixed and incubated with a rabbit polyclonal anti-LYVE-1 antibody (R&D Systems), washed with PBS and stained with a A488 conjugated anti-rabbit secondary antibody (Molecular Probes). The stained ears were mounted on custom-made migration chambers with the dermal surface exposed. LPS-matured BM-derived, MACS-purified, TAMRA-labelled DCs (see above) were re-suspended in culture medium, added on top of the dermis and incubated for 30 min at 37 °C, 5% CO<sub>2</sub>. After gently washing away non-infiltrated DCs, the ear sheets were imaged in a custom-built climate chamber using a fully motorized upright Leica MZ 16 FA stereomicroscope equipped with a Spot camera and operated via MetaMorph software (system implemented by Visitron Systems). For multiple immunofluorescence stainings (Supplementary Fig. 5), ear sheets were fixed with 1% PFA after 120 min immigration of DCs and stained as described in the extravasation section. Cells were manually tracked using ImageJ as described below.

**Two-photon intravital microscopy in lymph nodes.** Intravital microscopy was performed as described previously<sup>43</sup>. Briefly, BM-derived DCs were matured overnight with LPS, purified and selected by MACS for the desired integrin deficiency (see above), differentially labelled with TAMRA and CFSE and subcutaneously injected into footpads of C57BL/6 mice. After 32 h popliteal lymph nodes were surgically exposed and imaged using an upright Leica DM LFSa microscope equipped with a  $\times 20$  0.95 NA water immersion objective (Olympus) and a MaiTai Ti:sapphire-pulsed laser (Spectra-Physics). 3D cell tracking was performed using Imaris software (Bitplane) as described.

**In vivo ear irritation in bone marrow chimaeric mice.** BM chimaera of talin1<sup>fllox/flox</sup>, Mx1Cre<sup>+/-</sup> and  $\alpha_4$ <sup>fllox/flox</sup>,  $\beta_1$ <sup>fllox/flox</sup>,  $\beta_2$ <sup>-/-</sup>,  $\beta_2$ <sup>+/-</sup>, Mx1Cre<sup>+/-</sup> animals were generated by sublethal irradiation of wild-type mice and immediate reconstitution with  $1 \times 10^6$  suspended BM cells of the indicated genotype. Six weeks after irradiation Poly (I)·Poly (C) was injected and 10 d later 20  $\mu\text{l}$  of 1% croton oil (Sigma) in acetone<sup>44</sup> was applied on mouse ears. Mice were killed 12 h after application of croton oil. For histological analysis dorsal and ventral halves of ears were subjected to whole mount immunostaining. After fixation with 1% PFA, ear halves were blocked with 1% BSA (PAA) in PBS for 1 h at RT (while shaking), stained with anti-Gr-1 (RB6-8C5, Pharmingen) and anti-pan-laminin antibody (L9393, Sigma) in 1% BSA in PBS overnight at 4 °C (while shaking) and washed with 1% BSA in PBS. Primary antibodies were detected with a repeated cycle of staining with anti-rabbit A488 (Molecular Probes) and anti-rat Cy3 (Dianova) antibodies before tissue was embedded in elvanol and representative images taken with an Axio Imager (Zeiss) equipped with the Apotome. Concentration of granulocytes in blood was measured by staining peripheral blood with Gr-1 PE (BD Pharmingen) and counting all granulocytes per blood sample by flow cytometry using a FACSCalibur (BD Pharmingen). The numbers of granulocytes that extravasated into ear dermis were quantified by morphometric analysis employing MetaMorph imaging software (Molecular Devices).

**3D collagen gel chemotaxis assay.** For standard assays, PureCol (INAMED) in  $1 \times$  minimum essential medium eagle (MEM, Sigma) and 0.4% sodium bicarbonate (Sigma) was mixed with cells in RPMI (Invitrogen), 10% fetal calf serum (FCS, Invitrogen) at a 2:1 ratio, resulting in gels with a collagen concentration of 1.6 mg ml<sup>-1</sup>. Final cell concentrations in the assay were  $1 \times 10^6$  cells per ml gel for DCs and  $1.6 \times 10^6$  cells per ml gel for granulocytes and B lymphoblasts. For collagen titration experiments, above ratios were adapted. For gels with 3–5 mg ml<sup>-1</sup> collagen concentration, Nutragen (INAMED) was employed. Collagen-cell mixtures were cast in custom-made migration chambers with a thickness of 0.5–1 mm. After 30 min assembly of the collagen fibres at 37 °C, the gels were overlaid with 50  $\mu\text{l}$  of the following recombinant chemokines (all R&D Systems) diluted in RPMI, 10% FCS: CCL19 (0.6  $\mu\text{g}$  ml<sup>-1</sup>, DCs), C5a

(0.1  $\mu\text{g}$  ml<sup>-1</sup>, granulocytes) and CXCL13 (2.5  $\mu\text{g}$  ml<sup>-1</sup>, B cells). For inhibition studies, leukocytes were pre-incubated for 30 min with 50  $\mu\text{M}$  blebbistatin (Sigma), 30  $\mu\text{M}$  Y27632, 100 nM or 500 nM latrunculin B (both Calbiochem) before casting the gels. Inhibitors were also present in the same end-concentrations in the gel.

**3D fibrin gel chemotaxis assay.** Cells were taken up in 0.2% (w/v) human fibrinogen (IMCO) in RPMI, 10% FCS to a final concentration of  $1 \times 10^6$  cells per ml gel. Before casting the gels in custom-made migration chambers, 0.125 U ml<sup>-1</sup> thrombin (Calbiochem) was added. After 30 min polymerization of the fibrin gels at 37 °C, the gels were overlaid with 50  $\mu\text{l}$  of 0.6  $\mu\text{g}$  ml<sup>-1</sup> CCL19 (R&D Systems) in RPMI, 10% FCS.

**Analysis of chemotaxis assays.** For comparison of DC velocities in collagen and fibrin gel assays in the absence of inhibitors (Fig. 3b and Supplementary Fig. 7a), movies were converted to kymographs using MetaMorph imaging software. Straight lines in the kymographs represent migrating cells and lead angles were measured to calculate single cell velocities. Manual single cell tracking of random samples employing ImageJ and the Manual Tracking Plugin revealed the same results. Manual tracking was applied to all other 3D chemotaxis assays. Here cells were tracked over 4 h (for DCs, 2 min per frame) or 15 min (for granulocytes, 15 s per frame). Speed and directionality parameters were calculated and visualized as plots and movies by analysing the acquired data with the Chemotaxis and Migration Tool Plugin ([http://www.ibidi.de/applications/ap\\_chemo.html](http://www.ibidi.de/applications/ap_chemo.html)). For quantification of “resting” and “moving” phases in 0.75 mg ml<sup>-1</sup> collagen gels (Supplementary Fig. 9) instantaneous velocities (distance per 2 min) were categorized as “resting” if velocity < 1  $\mu\text{m}$  per 2 min and “moving” if velocity  $\geq 1$   $\mu\text{m}$  per 2 min. Average velocities during mobile phases were calculated by excluding the resting fraction. The percentage of resting phases was calculated as resting time / (resting time + moving time).

**Time-lapse video microscopy.** Low magnification bright-field movies of DCs in gels were recorded in indicated time intervals using inverted Axiovert 40 (Zeiss) cell culture microscopes, equipped with custom-built climate chambers (5% CO<sub>2</sub>, 37 °C, humidified) and PAL cameras (Prosilica) triggered by custom-made software (SVS Vistek). High magnification DIC and fluorescence movies were recorded using a fully automated Axiovert 200M inverted fluorescence microscope with a 40 $\times$  objective (Zeiss) and a CoolSNAP HQ CCD camera. The microscopic set-up (Visitron Systems) was controlled by MetaMorph software. **Transfection and labelling of cells.** DCs were transfected with MLC-GFP (gift of M. Olson<sup>45</sup>) between days 8–10 of culture using the primary mouse T cell kit and the Amara nucleoporator (Amara). Cells were transfected according to the manufacturer's recommendations and immediately after transfection LPS was added for overnight maturation. For nucleus visualization of DCs and granulocytes nucleic acid dyes SYTO11 and SYTO13 (Invitrogen) were used, respectively, according to the manufacturers recommendations.

**2D adhesion assays.** Adhesion assays were performed in 24-well plates coated with 10  $\mu\text{g}$  ml<sup>-1</sup> fibronectin (Calbiochem) or 10  $\mu\text{g}$  ml<sup>-1</sup> recombinant mouse ICAM-1 (R&D).  $3 \times 10^5$  DCs per well were seeded in RPMI medium containing 10% FCS in the presence of 200 ng ml<sup>-1</sup> LPS and incubated for 30 min at 5% CO<sub>2</sub>, 37 °C. Same cell numbers of LPS-matured DCs were seeded in the presence of 200 ng ml<sup>-1</sup> phorbol 12-myristate 13-acetate (Calbiochem) for 30 min. Cells were washed manually 3 times with PBS, 1% BSA and subsequently photo-documented and counted manually. Time-lapse video microscopy of chemokine-induced spreading of LPS-matured DCs (see Supplementary Video 4) was performed on 10  $\mu\text{g}$  ml<sup>-1</sup> mouse CCL21 (R&D)-coated plastic of cell culture dishes.

37. Bader, B. L., Rayburn, H., Crowley, D. & Hynes, R. O. Extensive vasculogenesis, angiogenesis, and organogenesis precede lethality in mice lacking all alpha v integrins. *Cell* **95**, 507–519 (1998).
38. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage Embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
39. Schaft, J., Ashery-Padan, R., van der Hoeven, F., Gruss, P. & Stewart, A. F. Efficient FLP recombination in mouse ES cells and oocytes. *Genesis* **31**, 6–10 (2001).
40. Bungartz, G. *et al.* Adult murine hematopoiesis can proceed without beta1 and beta7 integrins. *Blood* **108**, 1857–1864 (2006).
41. Lutz, M. B. *et al.* An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. *J. Immunol. Methods* **223**, 77–92 (1999).
42. Fässler, R. & Meyer, M. Consequences of lack of beta 1 integrin gene expression in mice. *Genes Dev.* **9**, 1896–1908 (1995).
43. Worbs, T., Mempel, T. R., Bolter, J., von Andrian, U. H. & Förster, R. CCR7 ligands stimulate the intranodal motility of T lymphocytes *in vivo*. *J. Exp. Med.* **204**, 489–495 (2007).
44. Berg, D. J. *et al.* Interleukin 10 but not interleukin 4 is a natural suppressant of cutaneous inflammatory responses. *J. Exp. Med.* **182**, 99–108 (1995).
45. Croft, D. R. *et al.* Actin-myosin-based contraction is responsible for apoptotic nuclear disintegration. *J. Cell Biol.* **168**, 245–255 (2005).

## ARTICLES

# Mapping and sequencing of structural variation from eight human genomes

Jeffrey M. Kidd<sup>1</sup>, Gregory M. Cooper<sup>1</sup>, William F. Donahue<sup>2</sup>, Hillary S. Hayden<sup>3</sup>, Nick Sampsas<sup>4</sup>, Tina Graves<sup>5</sup>, Nancy Hansen<sup>6</sup>, Brian Teague<sup>7</sup>, Can Alkan<sup>1</sup>, Francesca Antonacci<sup>1</sup>, Eric Haugen<sup>3</sup>, Troy Zerr<sup>1</sup>, N. Alice Yamada<sup>4</sup>, Peter Tsang<sup>4</sup>, Tera L. Newman<sup>1</sup>, Eray Tüzün<sup>1</sup>, Ze Cheng<sup>1</sup>, Heather M. Ebling<sup>2</sup>, Nadeem Tusneem<sup>2</sup>, Robert David<sup>2</sup>, Will Gillett<sup>3</sup>, Karen A. Phelps<sup>3</sup>, Molly Weaver<sup>1</sup>, David Saranga<sup>2</sup>, Adrienne Brand<sup>2</sup>, Wei Tao<sup>2</sup>, Erik Gustafson<sup>2</sup>, Kevin McKernan<sup>2</sup>, Lin Chen<sup>1</sup>, Maika Malig<sup>1</sup>, Joshua D. Smith<sup>1</sup>, Joshua M. Korn<sup>8</sup>, Steven A. McCarroll<sup>8</sup>, David A. Altshuler<sup>8</sup>, Daniel A. Peiffer<sup>9</sup>, Michael Dorschner<sup>1</sup>, John Stamatoyannopoulos<sup>1</sup>, David Schwartz<sup>7</sup>, Deborah A. Nickerson<sup>1</sup>, James C. Mullikin<sup>6</sup>, Richard K. Wilson<sup>5</sup>, Laurakay Bruhn<sup>4</sup>, Maynard V. Olson<sup>3</sup>, Rajinder Kaul<sup>3</sup>, Douglas R. Smith<sup>2</sup> & Evan E. Eichler<sup>1</sup>

Genetic variation among individual humans occurs on many different scales, ranging from gross alterations in the human karyotype to single nucleotide changes. Here we explore variation on an intermediate scale—particularly insertions, deletions and inversions affecting from a few thousand to a few million base pairs. We employed a clone-based method to interrogate this intermediate structural variation in eight individuals of diverse geographic ancestry. Our analysis provides a comprehensive overview of the normal pattern of structural variation present in these genomes, refining the location of 1,695 structural variants. We find that 50% were seen in more than one individual and that nearly half lay outside regions of the genome previously described as structurally variant. We discover 525 new insertion sequences that are not present in the human reference genome and show that many of these are variable in copy number between individuals. Complete sequencing of 261 structural variants reveals considerable locus complexity and provides insights into the different mutational processes that have shaped the human genome. These data provide the first high-resolution sequence map of human structural variation—a standard for genotyping platforms and a prelude to future individual genome sequencing projects.

Human genetic structural variation, including large (more than 1 kilobase pair (kbp)) insertions, deletions and inversions of DNA, is common<sup>1–9</sup>. These differences are thought to encompass more polymorphic base pairs than single nucleotide differences<sup>5,6,9,10</sup>. The importance of structural variation to human health and common genetic disease has become increasingly apparent<sup>11–14</sup>. However, only a small fraction of copy-number variant (CNV) base pairs have been determined at the sequence level<sup>15</sup>. Most genome-wide approaches for detecting CNVs are indirect, depending on signal intensity differences to predict regions of variation. They therefore provide limited positional information and cannot detect balanced events such as inversions. Because the human genome reference assembly is now viewed as a patchwork of structurally variant sequence<sup>1,2</sup>, it is expected that sequencing projects of other individuals would reveal previously uncharacterized human euchromatic sequence, in a similar manner to comparisons between the Celera and International Human Genome Project assemblies<sup>16–18</sup>. We implemented an approach to construct clone-based maps of eight human genomes with the aim of systematically cloning and sequencing structural variants more than 8 kbp in length. We present a validated structural variation map of these eight human genomes of Asian, European and African ancestry, identify 525 regions of previously

uncharacterized ‘novel sequence’, and provide sequence resolution of 261 selected regions of structural variation in the human genome.

## Fine-scale map of human genome structural variation

We selected eight individuals as part of the first phase of the Human Genome Structural Variation Project<sup>19</sup> (Supplementary Information). This included four individuals of Yoruba Nigerian ethnicity and four individuals of non-African ethnicity<sup>20</sup> (Table 1 and Supplementary Information). For each individual we constructed a whole genomic library of about 1 million clones by using a fosmid subcloning strategy<sup>21</sup>. Each library was arrayed and both ends of each clone insert were sequenced to generate a pair of high-quality end sequences (termed an end-sequence pair (ESP)<sup>22</sup>). The overall approach generated a physical clone map for each individual human genome, flagging regions discrepant by size or orientation on the basis of the placement of end sequences against the reference assembly (Supplementary Fig. 1)<sup>3,19</sup>. Across all eight libraries, we mapped 6.1 million clones to distinct locations against the reference sequence (Supplementary Fig. 2; <http://hgsv.washington.edu>). Of these, 76,767 were discordant by length and/or orientation (Supplementary Fig. 3 and Supplementary Table 1), indicating potential sites of structural variation. About 0.4% (23,742) of the ESPs mapped with only one

<sup>1</sup>Department of Genome Sciences and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Agencourt Bioscience Corporation, Beverly, Massachusetts 01915, USA. <sup>3</sup>Division of Medical Genetics, Department of Medicine, and University of Washington Genome Center, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>Agilent Technologies, Santa Clara, California 95051, USA. <sup>5</sup>Washington University Genome Sequencing Center, School of Medicine, St Louis, Missouri 63108, USA. <sup>6</sup>Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>7</sup>Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>8</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02114, USA. <sup>9</sup>Illumina, Inc., 9885 Towne Centre Drive, San Diego, California 92121, USA.



**Table 1 | Validated sites of structural variation detected by fosmid end sequence pairs**

Library information					Nucleotide variation		Structural variation					
Library	Population	Coriell ID	Number of reads*	Q30 bases†	Total SNPs‡	Total indels§	Mean <i>in silico</i> insert size   (kb)	Standard deviation¶ (kb)	Size threshold# (kb)	Deletions	Insertions	Inversions
G248		NA15510					39.89	2.75	8.25	105	108	48
ABC7	Yoruba	NA18517	2,076,237	841,505,234	705,558	125,307	37.59	3.88	11.64	86	42	59
ABC8	Yoruba	NA18507	3,331,676	1,549,030,580	1,241,616	222,740	36.7	3.85	11.55	156	101	67
ABC9	Japan	NA18956	2,076,828	1,032,726,952	610,897	114,203	39.51	2.26	6.78	163	186	79
ABC10	Yoruba	NA19240	2,118,546	1,022,392,331	736,451	130,646	41	1.84	5.52	252	297	88
ABC11	China	NA18555	1,966,644	939,700,332	590,518	102,650	40.03	1.77	5.31	286	246	67
ABC12	CEPH	NA12878	2,168,656	1,005,800,422	610,619	102,524	39.75	1.4	4.2	274	258	77
ABC13	Yoruba	NA19129	2,053,392	1,083,273,138	591,310	101,565	39.29	1.77	5.31	246	265	83
ABC14	CEPH	NA12156	2,021,844	1,086,415,113	714,848	140,498	39.44	1.72	5.16	275	265	98
Total			17,813,823	8,560,844,102	5,801,817	1,040,133				1,843	1,768	666
Non-redundant total					4,044,538	795,989				747	724	224

1,068 SNPs and 284 indels on the Y chromosome were identified from the single male sample (NA18507).

\* Number of sequencing reads generated from each library.

† Number of sequenced bases with a quality score of at least Q30 (99.9% accuracy).

‡ Single nucleotide variants with respect to the reference genome; no information regarding frequency.

§ Indels were insertion/deletion variants 1–100 bp in size.

|| Mean *in silico* insert size: mean size of clones based on mapping of paired reads.

¶ Standard deviation of *in silico* clone insert sizes based on ESP mapping.

# Size threshold (3 s.d.) used for detecting variant sites (see Supplementary Methods).

end to the reference assembly despite the presence of high-quality sequence at the other end (termed one-end anchored (OEA) clones; Supplementary Table 2 and Supplementary Information).

We undertook three main approaches to validate sites of copy-number variation. First, we selected 3,371 discordant fosmids corresponding to sites supported by two or more overlapping fosmids from the same individual whose apparent insert size deviated from the library mean insert size. These corresponded to 2,990 non-overlapping sites that are supported by multiple independent clones<sup>3</sup>. Using four multiple complete restriction enzyme digests (MCD analysis), we compared the predicted and expected insert sizes, confirming 1,182 non-redundant sites of copy-number variation (Supplementary Tables 3 and 4). As a secondary validation method, we designed two high-density customized oligonucleotide microarrays targeting a subset of insertion and deletion regions (Supplementary Fig. 4). This analysis recovered an additional 194 regions that had a copy-number difference but were not validated by MCD analysis. Combined with other experimental methods, we validated a total of 1,471 sites of copy-number variation (Fig. 1, Table 1, Supplementary Tables 3 and 4, and Supplementary Information). To assess the heritability of our events, we further intersected validated deletions with single nucleotide polymorphism (SNP) genotyping data (Illumina Human1M BeadChip) collected for 125 HapMap DNAs of African, European and Asian individuals, which included 28 parent–child trios. Although only a subset of the deletion events ( $n = 130$ ) could be reliably genotyped because of a lack of informative probes (Supplementary Fig. 5 and Supplementary Table 5), the allele frequencies ranged from rare (1%) to common (more than 50%), were generally consistent with Hardy–Weinberg equilibrium, and more than 98% of parent–child transmissions were consistent with mendelian patterns of inheritance (Supplementary Information).

Inversions proved more difficult to validate in a high-throughput manner because the events are balanced and because breakpoints are prone to map in the largest and most complex regions of segmental duplications<sup>23–25</sup>. We validated 217 inversions by detailed fingerprint analysis and/or sequence analysis. In addition, we validated seven larger ESP-detected inversions by interphase and metaphase fluorescence *in situ* hybridization (Supplementary Fig. 6, and Supplementary Tables 6 and 7). This included two previously described events: a roughly 5-million-base-pair (Mbp) inversion on 8p23.1 and a roughly 1-Mbp inversion on 17q21.3. We detected five novel large inversions, including a 1.2-Mbp inversion on 15q24, a 2.1-Mbp inversion on 15q13, and a 1.7-Mbp inversion on 17q12. Three of these regions correspond to sites of recurrent microdeletion associated with human disease, providing further support for a link between common

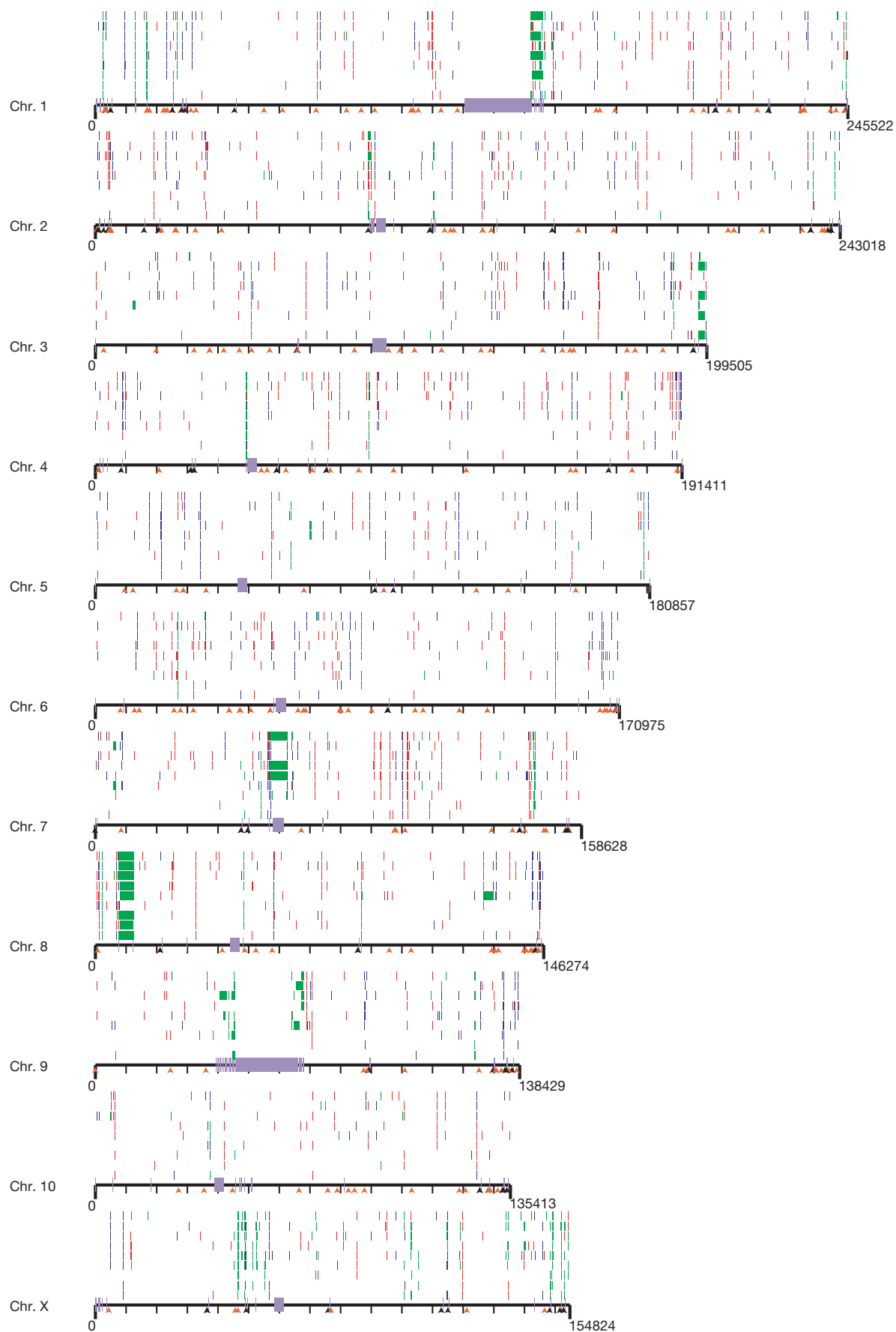
inversion polymorphisms and genomic disorders<sup>26,27</sup>. Overall, we found a twofold enrichment for inversions mapping to clustered regions of the X chromosome (Fig. 1 and Supplementary Table 7), consistent with theoretical predictions of increased inversion content based on unusual inverted repeat structures<sup>28</sup>. These data provide one of the first high-quality inversion maps of the human genome.

In total, we validated and refined the location of 1,695 sites of structural variation across nine diploid human genomes (eight fosmid libraries plus the original genome examined by the fosmid ESP approach (G248)) (Fig. 1, Table 1 and Supplementary Fig. 7). This included 747 deletions, 724 insertions and 224 inversions. A large fraction of the insertion/deletion events (40%) are novel when compared with previous published reports of CNVs. This is particularly unexpected, considering that at least 25% of the human genome now shows some evidence of copy-number variation (The Database of Genomic Variants<sup>1</sup>, hg17.v2). Many of the events (856, or 50%) were identified in multiple libraries and probably represent common polymorphisms (more than 5% frequency) (Fig. 2); 261 (15%) of the sites were observed in five or more individuals, indicating that the current reference human genome sequence organization may actually represent a minor allele. At 34 loci, all nine individuals were inconsistent with the build35 assembly, identifying the reference allele as rare or as a potential sequence misassembly.

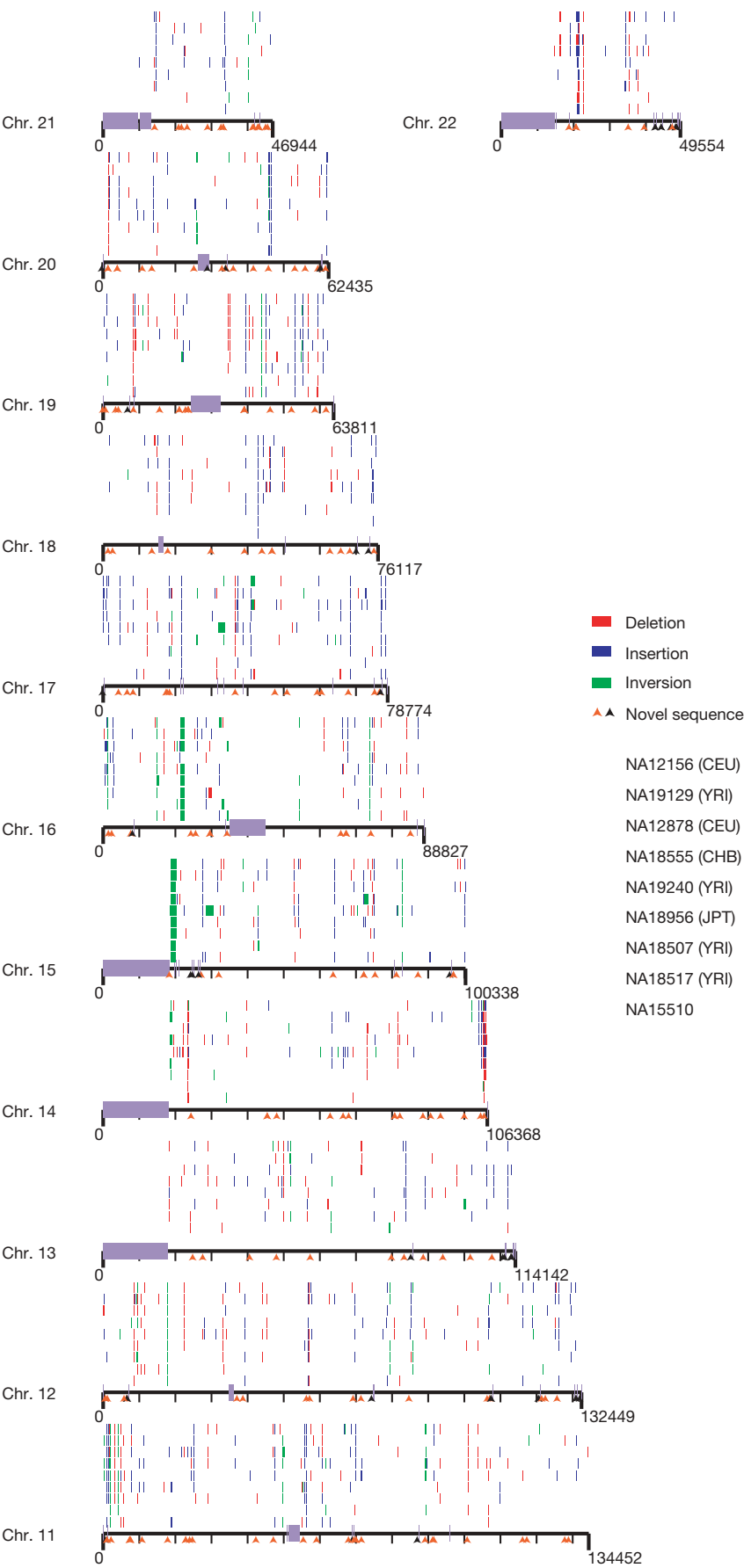
Using the refined set of CNVs, we compared CNV predictions within eight of the same samples analysed in ref. 5 (Supplementary Information). When we compared the predicted size of intersected sites on the same eight samples, we found that the bacterial artificial chromosome (BAC) array comparative genomic hybridization (CGH) CNVs were substantially (tenfold) larger and showed no correlation with the ESP estimated size (Supplementary Fig. 8). In contrast, we found extremely strong concordance between the sizes estimated from the ESP map and the annotations generated by our targeted high-density array CGH experiments (Supplementary Fig. 8b) and independent predictions on the same eight individuals analysed using the Affymetrix 6.0 platform (Supplementary Information and Supplementary Fig. 8c). We conclude that the BAC array CGH experiments performed in ref. 5 had, in some cases, exquisite sensitivity to detect much smaller events (about 10 kbp) than previously expected. However, our analysis indicates that the current amount of the reference genome sequence represented as CNV in these eight genomes has been overestimated.

### Novel human euchromatic sequences

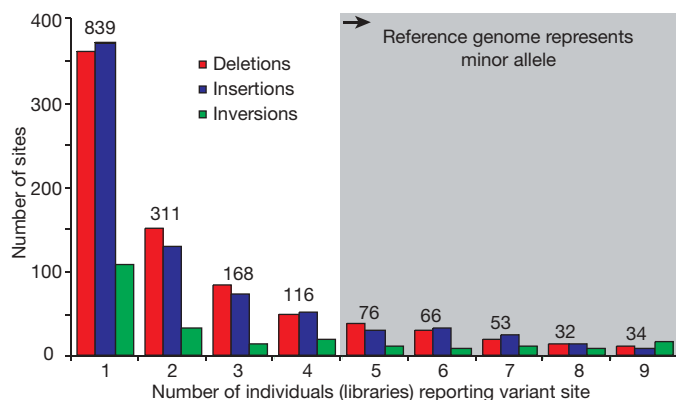
To identify potentially novel euchromatic sequences not present within the reference genome, we first identified clusters of clones







**Figure 1 | Map of structural variation in the human genome.** The location of 724 insertions (blue), 747 deletions (red) and 224 inversions (green) that have been experimentally validated are mapped onto the human genome (build35). Sites are arranged according to individuals in rows above each chromosome, in order of the nine individual genomic libraries (G248 (first row), then ABC7–ABC14); the Coriell IDs are listed in Table 1. All sites have been validated by array CGH, MCD analysis, or sequencing in at least one reference individual. The location of 525 novel sequence loci are depicted as arrows below each chromosome. Those mapping to gaps (black) are distinguished from those mapping to regions not associated with gaps (orange). The Y chromosome is not shown because samples were primarily from females.



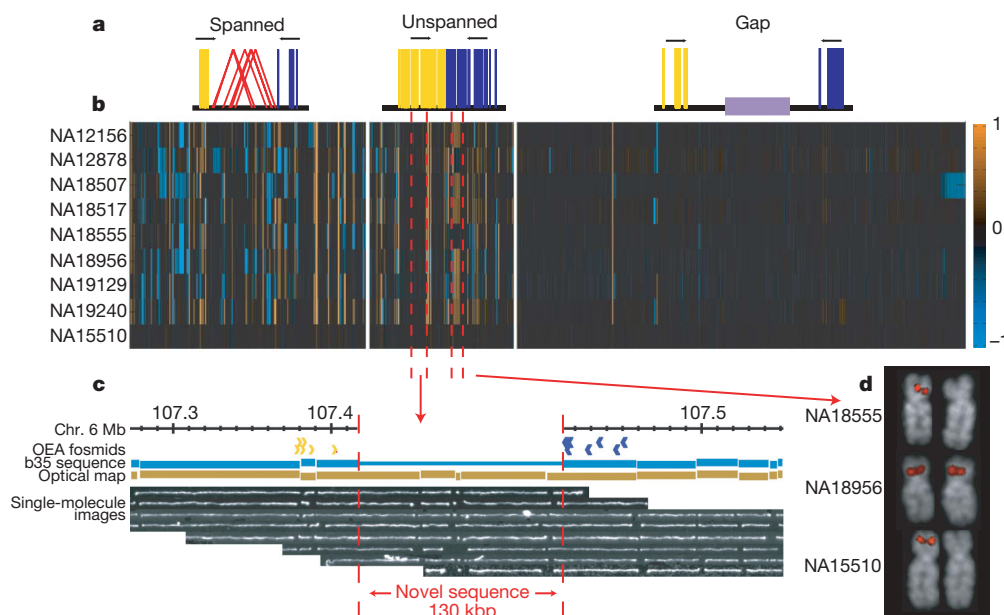
**Figure 2 | Frequency distribution.** Plot showing the number of times that a particular structural variant was detected on the basis of ESP analysis for nine fosmid libraries (eight HapMap, plus G248): 15% (261 of 1,695) of the sites seem to represent a more common sequence configuration (major allele) with respect to the human reference genome; 49% (839 of 1,695) of the validated sites are observed once, suggesting that saturation has not been achieved. The numbers above the columns report the total number of events for each frequency class.

in which one end sequence mapped to the human reference assembly but the other end sequence did not, termed OEA clusters (Fig. 3a and Supplementary Information). Pooling results from the first seven genomic libraries, we identified 21,556 OEA clones. Next, we assembled the sequence corresponding to all non-anchored ends by using the TIGR assembler<sup>29</sup>. This procedure generated 1,736 sequence contigs ( $n = 4,996$  OEA clones) of which 48% (820) had

no matches to previously published human sequence assemblies (minimum 100 base pairs (bp) with more than 98% sequence identity). By combining these sequence contigs with the positions of the OEA clusters we identified the map location of 525 regions of novel sequence insertion.

We distinguished three categories of novel insertion (Fig. 3a and Supplementary Fig. 9): 214 of the novel insertion loci intersected with regions identified as insertions with the paired-end sequence approach (see above); 139 putative 'insertions' flanked sequence assembly gaps<sup>30</sup>; and another 172 new sites did not correspond to known gaps or spanned insertions within the human genome. Among these we identified at least 11 regions where we estimate that the insertions are too large (more than 40 kbp in length) to be physically spanned by fosmid ESPs. Examination of these loci in a whole-genome restriction map constructed by optical mapping (Supplementary Information) on one of the same individuals confirms that the majority (8 of 11) correspond to insertions as large as 130 kbp in length (Fig. 3c and Supplementary Table 8).

To assess copy-number variation of these unannotated human sequences, we designed an oligonucleotide microarray specifically for these 525 loci and assessed copy-number status by array CGH (Supplementary Information) among the eight genomes tested (Fig. 3b). Novel sequences not associated with gaps showed the most extensive variation in copy number. For example, we found that 49% of novel sequences associated with fosmid ESPs (spanned insertions) showed evidence of copy-number variation. We note that sequence contigs mapping to the same novel locus (Fig. 3d) often showed the same pattern of copy-number variation. Such regions cannot be genotyped by existing commercial platforms that depend on sequence in the reference genome. The presence of a mapped clone



**Figure 3 | Discovery of novel human sequences that are CNV.** **a**, Clusters of clones where one end is mapped to the genome (build35) but the other does not map are shown schematically on the basis of their orientation (blue and yellow lines). Three categories are distinguished: clones mapping around a site already spanned by a discordant fosmid ESP (spanned), regions where no discordant clones are identified (unspanned), and clones mapping adjacent to sequence gaps (gap). **b**, Array CGH experiment based on an oligonucleotide microarray designed to a sequence assembly of these novel sequences (525 distinct loci). Of the spanned and unspanned loci, 45% show copy-number variation (gains, orange; losses, blue) in comparison with a reference sample (NA15510). Each data point represents the average  $\log_2$  intensity values for all of the probes from a single contig. Within each of the three categories, contigs are ordered on the basis of their chromosomal anchored positions. The bottom row represents the results of one of three

self-versus-self hybridizations with sample NA15510. **c**, A novel insertion of 130 kbp on chromosome 6 identified by OEA fosmid clones (blue and gold arrows) and confirmed by optical mapping of DNA from the GM15510 cell line. Optical images of *Swa*I-restricted DNA are aligned to the reference (build35) genome. This large insertion maps intergenically to a region rich in conserved sequence elements and is confirmed in all eight libraries. This region does not correspond to a known gap in the human genome and does not appear CNV in our eight samples. **d**, Validation of a CNV region by fluorescence *in situ* hybridization. Hemizygous signals are detected by fluorescence *in situ* hybridization on metaphase chromosomes (with OEA clone ABC7\_42397600\_G7 as probe), corresponding to samples where no signal intensity difference was observed with respect to the reference by array CGH.



ensures that these regions can be sequenced in their entirety (see below) and incorporated as part of future CNV and SNP genotyping platforms.

### Sequence resolution

The acquisition of high-quality, finished sequence corresponding to the breakpoints of a rearrangement is the ultimate form of validation<sup>31</sup>. We selected 405 clones with predicted structurally variant haplotypes by ESP analysis for full-length sequencing (Supplementary Fig. 10 and Supplementary Table 9), generating 16 Mbp of alternative haplotype sequence. Sequence validation confirmed 230 insertion/deletion loci (median 8.1 kbp, mean 15.3 kbp) and 35 inversions (up to 2 Mbp in size). Validation for 63 sequenced clones could not be conclusively resolved, despite the fact that both fingerprint and ESP analysis confirmed 80% of these 'ambiguous' clones as being structurally variant with respect to the reference genome. Detailed sequence analyses revealed that most of these contained large, multi-copy tandem repeat sequences, which confound breakpoint identification and complicate the final sequence assembly of the insert. Including these ambiguous clones, we estimate that 84% (341 of 405) of the clones contained structural variants. The vast majority of the clones that failed to confirm at the sequence level represented putative insertion events (Supplementary Information) as a result of a slight subcloning preference for 'short insert' clones.

High-quality finished sequence at the breakpoints allowed us to assess the potential molecular mechanisms underlying larger structural variation events in the human genome (Table 2). Non-allelic homologous recombination between repeated sequences accounts for 47% (124 of 261) of events assigned a mechanism. Recombination between segmental duplications is more common than L1 or Alu-mediated events. Of the inversions, 67% show evidence of large blocks of sequence homology at the breakpoints, with the remainder mediated by shorter common repeat sequences. An additional ten events (4%) involved the expansion or contraction of a variable number of tandem repeats. Retrotransposition accounted for 15% (40 of 261) of events, although this is likely to represent a lower bound given that the detection thresholds exceeded the length of an L1 insertion (6 kbp) for several of the libraries (Table 1). Analysis of structurally variant sequences found a slight enrichment of repetitive DNA for both insertion (58.5%) and deletion (60.8%) events, with 28% of events having a repeat content greater than 90%. Such events are not resolvable with array-based techniques and will probably require directed, PCR-based assays for genotyping.

We compared RefSeq gene annotation between the structurally variant haplotypes and found that 107 distinct gene structures were altered (Supplementary Table 10). Of these genes, 87% belong to members of a gene family, suggesting potential functional redundancy. We specifically examined insertion sequences and found homology for 60 spliced expressed sequence tags and 15 RefSeq gene annotations. Most of these putative gene structures corresponded to duplicated copies of genes or portions of genes (*NAIP*, *BIRCA1*, *NBPF11*, *DNM1* and *LPA*) and/or had homology to genes predicted in either chimpanzee or macaque (*ANKRD20A* and *LOC713531*). There are three examples of insertions restricted to coding exons (*EPPK1*, *BAHCC1* and *MUC6*)—events predicted to alter the composition and structure of the encoded transcripts and proteins. In the case of *MUC6* and *LPA*, these protein length polymorphisms have

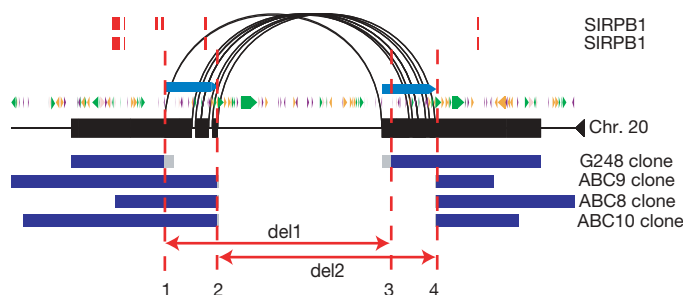
been associated with *H. pylori* infection<sup>32</sup> and risk of coronary heart disease<sup>33</sup>, respectively.

We sequenced multiple alleles for the *SIRPB1* locus and found evidence for recurrent deletion events on different haplotypes. Sequencing confirmed two distinct deletion alleles having different breakpoints (Fig. 4) embedded within segmental duplications. Both deletion alleles seem to be common and only one of these two results in the loss of an exon, raising the possibility that the two events have different functional consequences despite their extensive overlap. The two different alleles cannot be reliably distinguished by array CGH genotyping because of the presence of duplicated sequences at the boundary and uncertainty in the reference sample genotype (Supplementary Fig. 4). Although we have only begun to survey the sequence organization of a small fraction of our sites, a preliminary analysis of the SNP content of sequenced sites suggests that about 24% of the variants predicted in multiple individuals may be found on different haplotype backgrounds.

### Other forms of genetic variation

One of the ancillary benefits of sequence-based detection of structural variation is the identification and characterization of other forms of human genetic variation. Because each library represents about 0.3-fold sequence coverage per individual, the ESP pipeline generated about three genomic equivalents ( $8.5 \times 10^9$  bases) of high-quality sequence data from the eight individuals. We mined the existing 13 million end sequences<sup>34</sup> and identified 4.0 million non-redundant single nucleotide variants and 796,273 smaller insertion/deletion events (more than 1 bp to less than 100 bp in length); 28% (1.29 million) of the single nucleotide variants and 75% (597,790) of the insertion/deletion variants (indels) were novel when compared with dbSNP (build125). Of the eight HapMap individuals selected in this project, five are common to the ENCODE resequencing project<sup>35</sup>. We therefore compared our SNP and indel predictions against those ten regions resequenced in the same individuals as a measure of SNP/indel accuracy. On the basis of 1,988 SNP and 120 indel genotypes, we estimated false positive rates of 3.5% (SNPs) and 10.0% (indels).

As expected, the Yoruba African samples showed 15.3% more single nucleotide genetic diversity than non-African samples on the autosomes. The X chromosome shows greater genetic diversity (40%) between African and non-African samples when compared with the autosomes. Because this is one of the first random surveys of sequence data from an ethnically diverse collection of individuals,



**Figure 4 | Sequence resolution of human structural variation.** Two different deletions within the *SIRPB1* gene (exons, red) provide evidence for an independently recurrent deletion event. Both structural variants are probably mediated by non-allelic homologous recombination between segmental duplications (blue bars, arching lines) in direct orientation. Deletion alleles from four different individuals (G248 and ABC8–10) are depicted; deletion 2 (del2, minimal region chromosome 20: 1509210–1542041) eliminates exon 2, whereas deletion 1 (del1, minimal region chromosome 20: 1502353–1533914) does not. Repeat content and orientation are depicted as coloured arrows (green, long interspersed transposable element; purple, short interspersed transposable element; orange, transposon). Predicted and annotated segmental duplications are depicted as indicated.

**Table 2 | Inferred mechanism of sequenced structural variants**

Event type	Total events	NAHR	NHEJ	VNTR	Retrotransposition
Insertion	98	41	29	8	20
Deletion	129	49	58	2	20
Inversion	34	34	0	0	0
Total	261	124	87	10	40

The mechanisms of origin for 261 events were classified as being non-allelic homologous recombination (NAHR), non-homologous end-joining (NHEJ), variable number of tandem repeats (VNTR) or retrotransposition.

we also assessed single nucleotide density within 100-kbp windows across the entire genome, identifying regions significantly enriched or depleted in single nucleotide variants (Supplementary Information). After masking sites of segmental duplication, we identified 15 large regions of excess nucleotide variation, ranging in size from 500 kbp to 3 Mbp. These include known sites of increased sequence diversity<sup>36,37</sup>, for instance HLA and 8p23, as well as several previously undescribed regions such as two large (more than 10 Mbp) regions on each arm of chromosome 16 (Fig. 5). The interval on 8p23 also showed the highest concentration of structural variants validated by our ESP approach (22 distinct variants). The molecular basis for this regional enrichment of genetic diversity across human genomes is unknown, but our preliminary data suggest that structural and single nucleotide variation may correlate.

## Discussion

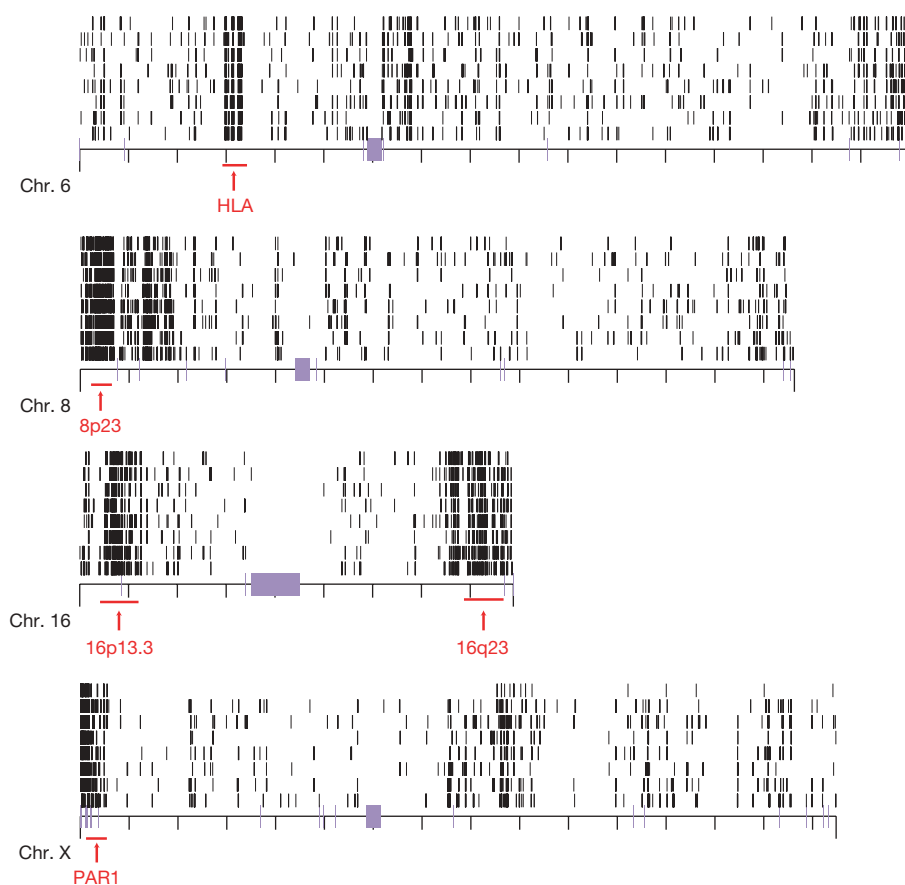
We present a high-resolution integrated map of genetic variation for eight human genomes. We refine the location of 1,695 sites of structural variation (more than about 6 kbp in length), identify 525 regions of novel sequence that harbour highly polymorphic CNVs, and provide single-base-pair sequence resolution for 261 regions of structural variation. These events are placed within the context of 4 million SNPs and 796,273 small indels (1–100 bp in size).

Our detailed analysis of eight human genomes provides significant biological and technological insights into human genetic variation. First, we have discovered and mapped a large number of novel sequences not represented in the human reference genome and show that more than 40% are CNV. These sequences range in size from a few kilobase pairs up to 130 kbp and are randomly distributed, located both within genic and intergenic regions. Although the

sequences represent only a fraction of the euchromatin (less than 0.1%), these results strongly argue that the human genome sequence is still incomplete. The role of such sequences in disease association cannot be determined without *de novo* sequencing of additional genomes and the design of new platforms to genotype these variants specifically on the basis of these 'new' sequences.

Second, our refined map of structural variation predicts that the current database of copy-number variation is inflated, which is consistent with previous studies<sup>38</sup>. An analysis of the same samples with customized high-resolution microarrays and two independent commercial platforms shows an excellent correspondence between ESP-predicted size and commercial SNP platforms (Affymetrix 6.0 arrays and Illumina Human1M BeadChips). The net effect is that there are fewer CNV base pairs per haplotype; consequently, fewer genes and exons are affected. This is an important consideration in view of the fact that CNV maps and databases based largely on BAC-based array CGH are being used to exclude disease-causing variation<sup>26,27,39</sup>. A comparison of the same eight individuals with the highest-density SNP commercial platforms reveals that more than 50% of the structural variants that we have detected cannot be adequately genotyped, although we note that many more events can be detected than is possible with the fosmid ESP approach. These data argue for the need for customized CNV genotyping platforms based on sequence-validated sites of structural variation.

Third, our sequence analyses suggest that non-allelic homologous recombination is the predominant mechanism (48%; Supplementary Information) altering the larger structural variation landscape of the human genome. This is consistent with several reports confirming that copy-number variation is enriched fourfold to tenfold for regions of segmental duplication<sup>3–5</sup>. However, these findings



**Figure 5 | Regions of enriched SNP density.** Regions of increased single nucleotide variant density identified in eight individuals are shown (ABC7–ABC14 samples ordered from bottom to top). Heterozygosity was calculated in 100-kbp windows, and those windows having a heterozygosity

2 s.d. above the mean are plotted for four chromosomes. Regions of increased heterozygosity, over 1 Mb of genome sequence, are highlighted by red bars. Purple bars, centromeres.



are in contrast with a recent analysis of two individuals with next-generation ESP sequencing technology<sup>40</sup>, which reported that events mediated by non-allelic homologous recombination were relatively rare. One possibility for this discrepancy is that shorter reads with lower sequence quality offered by next-generation sequencing technologies may have less power to map within duplication and repeat-rich regions of the genome, thereby missing a large fraction of variation.

The establishment of a clone-based framework for each of these eight genomes provides an important resource for future studies of genetic variation. The clones provide the ability to recover and integrate all forms of genetic variation, ranging from SNPs to larger structural variants within specific haplotypes. The existing end-sequenced clone map permits novel insert sequences anchored within the genome to be mapped and sequenced completely, generating complete alternative human haplotypes. It also permits sequence-based validation of CNV events predicted with other methods and facilitates the targeted resequencing of any genomic region of interest. These full insert sequences will be important in the identification of haplotype-specific 'tag' SNPs that may be used to genotype more complex structural variants indirectly and to assess more fully the spectrum of human genetic variation. Thus, these eight genomes can serve as an important benchmark as new genomes become routinely sequenced with next-generation technologies.

## METHODS SUMMARY

**Library construction and ESP analysis.** Fosmid libraries (pCC2Fos vector) were constructed<sup>21</sup> from human genomic DNA samples (Coriell Cell Repositories) corresponding to eight HapMap individuals (Table 1). We sequenced about 1 million clones (900 Mbp) for each genome in the form of high-quality ESPs (Supplementary Table 11) and deposited sequences into the NIH trace repository (<http://www.ncbi.nlm.nih.gov/Traces>). All ESPs were mapped to the human genome assembly (build35) with a previously described algorithm<sup>3</sup>. Map information, including ESP alignments and corresponding clone IDs of discordant and concordant clones, are available in an interactive browser format and database (<http://hgsv.washington.edu>).

**Validation.** Fosmid clones discordant by size ( $n = 3,371$  fosmid clones) were subjected to fingerprint analysis using four multiple complete restriction enzyme digests (MCD analysis) to confirm insert size and eliminate rearranged clones<sup>41,42</sup>. Two high-density customized oligonucleotide microarrays (Agilent and NimbleGen) were designed to confirm sites of deletion and insertion (GEO accessions GSE10008 and GSE10037). We developed a new, expectation maximization-based clustering approach to genotype deletions with the use of data from the Illumina Human1M BeadChip collected for 125 HapMap DNA samples (Supplementary Information). We found that more than 98% of the children's genotypes were consistent with mendelian transmission on the basis of an analysis of 28 parent-child trios.

**Fosmid insert sequencing.** We completely sequenced the inserts of 405 fosmid clones from six genomic libraries (210 from G248, 31 from ABC7, 39 from ABC8, 21 from ABC9, 98 from ABC10, and 6 from ABC12) with previously described methods. All sequences have been deposited in GenBank (Supplementary Table 9).

**SNP/indel analysis.** We identified single nucleotide variants using the ssahaSNP software tool<sup>34</sup> and indels (1–100 bp in size) using ssahaSNP in combination with cross\_match (<http://www.phrap.org>; Supplementary Information). SNP and indel variants have been deposited in dbSNP (release 129).

A detailed materials and methods section can be found as part of the Supplementary Information.

Received 7 November 2007; accepted 15 February 2008.

1. Lafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
3. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
4. Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
5. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
6. Wong, K. K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80**, 91–104 (2007).
7. Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphisms in the human genome. *Nature Genet.* **38**, 75–81 (2006).
8. McCarroll, S. A. *et al.* Common deletion polymorphisms in the human genome. *Nature Genet.* **38**, 86–92 (2006).
9. Hinds, D. A., Kloek, A. P., Jen, M., Chen, X. & Frazer, K. A. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genet.* **38**, 82–85 (2006).
10. Cheng, Z. *et al.* A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**, 88–93 (2005).
11. Aitman, T. J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
12. Gonzalez, E. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
13. Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human  $\beta$ -defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.* **79**, 439–448 (2006).
14. Hollox, E. J. *et al.* Psoriasis is associated with increased  $\beta$ -defensin genomic copy number. *Nature Genet.* **40**, 23–25 (2007).
15. Cooper, G. M., Nickerson, D. A. & Eichler, E. E. Mutational and selective effects on copy-number variants in the human genome. *Nature Genet.* **39**, S22–S29 (2007).
16. Istrail, S. *et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl Acad. Sci. USA* **101**, 1916–1921 (2004).
17. Khaja, R. *et al.* Genome assembly comparison identifies structural variants in the human genome. *Nature Genet.* **38**, 1413–1418 (2006).
18. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
19. Eichler, E. E. *et al.* Completing the map of human genetic variation. *Nature* **447**, 161–165 (2007).
20. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
21. Donahue, W. & Ebling, H. M. Fosmid libraries for genomic structural variation detection. *Curr. Protocols Hum. Genet.* **5**, 20.1–20.18 (2007).
22. Volik, S. *et al.* End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc. Natl Acad. Sci. USA* **100**, 7696–7701 (2003).
23. Small, K., Iber, J. & Warren, S. Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nature Genet.* **16**, 96–99 (1997).
24. Giglio, S. *et al.* Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. *Am. J. Hum. Genet.* **71**, 276–285 (2002).
25. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37**, 129–137 (2005).
26. Sharp, A. J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature Genet.* **38**, 1038–1042 (2006).
27. Sharp, A. J. *et al.* Characterization of a recurrent 15q24 microdeletion syndrome. *Hum. Mol. Genet.* **16**, 567–572 (2007).
28. Warburton, P. E., Giordano, J., Cheung, F., Gelfand, Y. & Benson, G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* **14**, 1861–1869 (2004).
29. Sutton, G. G., White, O., Adams, M. D. & Kerlavage, A. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci. Technol.* **1**, 9–19 (1995).
30. Bovee, D. *et al.* Closing gaps in the human genome with fosmid resources generated from multiple individuals. *Nature Genet.* **40**, 96–101 (2008).
31. Scherer, S. W. *et al.* Challenges and standards in integrating surveys of structural variation. *Nature Genet.* **39**, S7–S15 (2007).
32. Nguyen, T. V. *et al.* Short mucin 6 alleles are associated with *H. pylori* infection. *World J. Gastroenterol.* **12**, 6021–6025 (2006).
33. Lackner, C., Cohen, J. C. & Hobbs, H. H. Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum. Mol. Genet.* **2**, 933–940 (1993).
34. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
35. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
36. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
37. Nusbaum, C. *et al.* DNA sequence and analysis of human chromosome 8. *Nature* **439**, 331–335 (2006).
38. de Smith, A. J. *et al.* Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum. Mol. Genet.* **16**, 2783–2794 (2007).
39. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
40. Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
41. Gillett, W. *et al.* Assembly of high-resolution restriction maps based on multiple complete digests of a redundant set of overlapping clones. *Genomics* **33**, 389–408 (1996).

42. Wong, G. K., Yu, J., Thayer, E. C. & Olson, M. V. Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. *Proc. Natl Acad. Sci. USA* **94**, 5225–5230 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the staff from the University of Washington Genome Center and the Washington University Genome Sequencing Center for technical assistance. J.M.K. is supported by a National Science Foundation Graduate Research Fellowship. G.M.C. is supported by a Merck, Jane Coffin Childs Memorial Fund Postdoctoral Fellowship. This work was supported by National Institutes of Health grants HG004120 to E.E.E., D.A.N. and M.V.O., and 3 U54 HG002043 to M.V.O. E.E.E. is an Investigator of the Howard Hughes Medical Institute.

**Author Contributions** J.M.K., G.M.C., M.V.O., D.A.N. and E.E.E. contributed to the writing of this paper. The study was coordinated by L.B., M.V.O., R.K., D.R.S., J.M.K. and E.E.E. A.B., D.R.S., D.Sa., E.G., H.M.E., K.M., N.T., R.D., W.F.D. and W.T. performed library construction and end sequencing. E.H., H.S.H., K.A.P., M.V.O., R.K., R.K.W., T.G. and W.G. performed clone insert validation and sequencing. C.A., D.A.N., E.T., J.D.S., J.S., L.C., M.D., M.M., M.W., T.L.N. and Z.C. provided technical and analytical support. D.A.P., D.A.A., J.M.Ko. and S.A.M. contributed variation data. G.M.C., J.M.K., L.B., N.A.Y., N.S. and P.T. designed and analysed array CGH experiments. G.M.C. and T.Z. performed the genotype analysis. F.A. performed FISH experiments. B.T. and D.S. performed optical mapping experiments. E.E.E., J.M.K. and L.C. analysed sequenced clones. J.C.M. and N.H. identified SNPs and indels.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the paper on [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to E.E.E. ([eee@gs.washington.edu](mailto:eee@gs.washington.edu)).



# Control of T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation by the aryl hydrocarbon receptor

Francisco J. Quintana<sup>1</sup>, Alexandre S. Basso<sup>1</sup>, Antonio H. Iglesias<sup>1</sup>, Thomas Korn<sup>1</sup>, Mauricio F. Farez<sup>1</sup>, Estelle Bettelli<sup>1</sup>, Mario Caccamo<sup>2</sup>, Mohamed Oukka<sup>3</sup> & Howard L. Weiner<sup>1</sup>

**Regulatory T cells (T<sub>reg</sub>) expressing the transcription factor Foxp3 control the autoreactive components of the immune system. The development of T<sub>reg</sub> cells is reciprocally related to that of pro-inflammatory T cells producing interleukin-17 (T<sub>H</sub>17). Although T<sub>reg</sub> cell dysfunction and/or T<sub>H</sub>17 cell dysregulation are thought to contribute to the development of autoimmune disorders, little is known about the physiological pathways that control the generation of these cell lineages. Here we report the identification of the ligand-activated transcription factor aryl hydrocarbon receptor (AHR) as a regulator of T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation in mice. AHR activation by its ligand 2,3,7,8-tetrachlorodibenzo-p-dioxin induced functional T<sub>reg</sub> cells that suppressed experimental autoimmune encephalomyelitis. On the other hand, AHR activation by 6-formylindolo[3,2-b]carbazole interfered with T<sub>reg</sub> cell development, boosted T<sub>H</sub>17 cell differentiation and increased the severity of experimental autoimmune encephalomyelitis in mice. Thus, AHR regulates both T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation in a ligand-specific fashion, constituting a unique target for therapeutic immunomodulation.**

Pathogenic autoimmunity is controlled in healthy individuals by a specialized subset of T cells named T<sub>reg</sub> cells<sup>1</sup>. T<sub>reg</sub> cell differentiation and function are driven by the transcription factor Foxp3 (refs 2, 3). Mice carrying a mutant Foxp3 show impaired T<sub>reg</sub> cell activity and succumb to a fatal lymphoproliferative disorder, which can be prevented by the transgenic expression of wild-type Foxp3 (ref. 4). FOXP3 is also important for human T<sub>reg</sub> cells; mutations in FOXP3 have been linked to the autoimmune syndrome immune dysregulation, polyendocrinopathy, enteropathy X-linked (IPEX)<sup>5,6</sup>.

T<sub>reg</sub> cell development is closely related to the generation of interleukin (IL)-17-producing T cells (T<sub>H</sub>17). Transforming growth factor (TGF)-β1 induces the differentiation of T<sub>reg</sub> cells<sup>7</sup>, whereas TGF-β1 in combination with IL-6 (ref. 8) or IL-21 (ref. 9) results in the differentiation of T<sub>H</sub>17 cells. T<sub>H</sub>17 cells express the transcription factor ROR-γt<sup>10</sup>, participate in the control of extracellular pathogens and have an important role in human and experimental autoimmunity<sup>11</sup>. Thus, the identification of the pathways that control T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation is of importance for the treatment of autoimmunity, infections and cancer.

Here we report that the aryl hydrocarbon receptor (AHR) regulates the generation of T<sub>reg</sub> and T<sub>H</sub>17 cells in mice. AHR activation by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) induced T<sub>reg</sub> cells that suppressed experimental autoimmune encephalomyelitis (EAE) by a TGF-β1-dependent mechanism, whereas AHR activation by 6-formylindolo[3,2-b]carbazole (FICZ) interfered with T<sub>reg</sub> cell differentiation, boosted T<sub>H</sub>17 cell differentiation and worsened EAE. Thus, AHR regulates T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation in a ligand-specific manner.

## AHR controls T<sub>reg</sub> cell generation

Our interest in the zebrafish as an immunological model led us to clone and characterize the zebrafish functional homologue of Foxp3. The expression of zebrafish Foxp3 resembled that of mammalian Foxp3 in that it was restricted to lymphocytes, suggesting the

conservation of mechanisms controlling tissue-specific Foxp3 expression. To identify evolutionarily conserved signalling pathways controlling Foxp3 expression, we performed a phylogenetic footprinting analysis to define conserved regulatory regions within the zebrafish, mouse and human Foxp3 gene<sup>12</sup>. We identified an evolutionarily conserved binding site for AHR (conserved AHR-binding site, CABS); a similarly located regulatory sequence controls the expression of the AHR-regulated gene cytochrome P4501A2 (*Cyp1a2*)<sup>13</sup> (Fig. 1a and Supplementary Fig. 1a, b). In addition, we identified three non-evolutionarily conserved AHR-binding sites (NCABS) in the Foxp3 promoter (termed NCABS-1, -2 and -3) where functional binding sites for other transcription factors have been found previously<sup>14–16</sup> (Supplementary Fig. 1a, c).

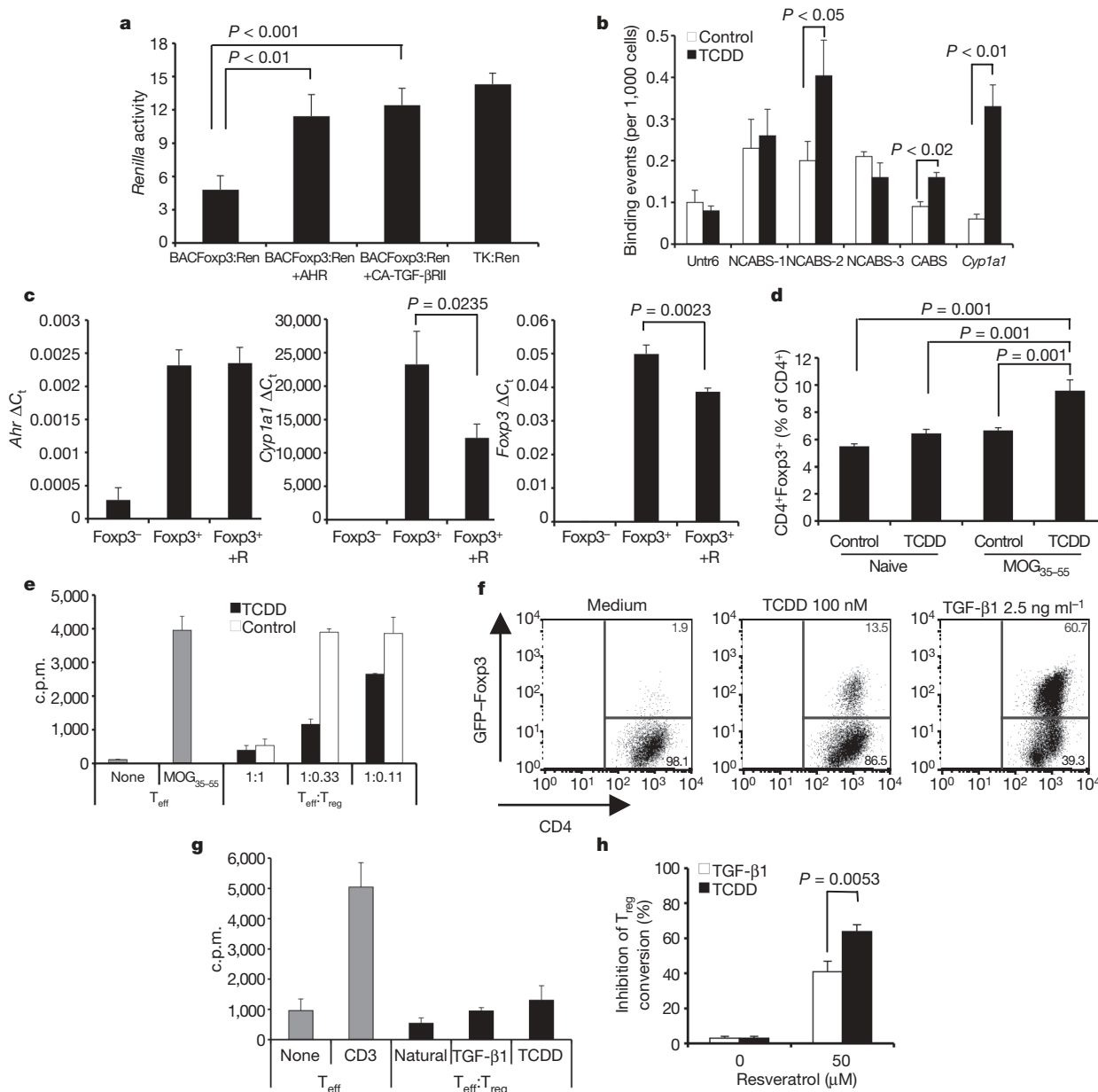
To investigate whether AHR directly controls Foxp3 expression we used a bacterial artificial chromosome that contained the entire mouse Foxp3 locus tagged with a *Renilla* luciferase reporter after the ATG start codon (BACFoxp3:Ren). Co-transfection of BACFoxp3:Ren with a construct coding for mouse AHR resulted in a significant upregulation of *Renilla* activity ( $P < 0.01$ ), similar to that achieved with constitutively activated TGF-β receptor II (Fig. 1a). We used chromatin immunoprecipitation (ChIP) to analyse the interaction of AHR with the CABS and NCABS shown in Fig. 1a. Treatment of CD4<sup>+</sup> T cells with TCDD increased AHR binding to the CABS and NCABS-2 ( $P < 0.05$ , Fig. 1b). This upregulation was comparable to that detected in the promoter of the AHR-regulated gene cytochrome P4501A1 (*Cyp1a1*) (Fig. 1b)<sup>17</sup>. No significant increase in AHR binding was seen in NCABS-1, NCABS-3 or the control sequence Untr6. Similar results were obtained when CD4<sup>+</sup> T cells were purified from TCDD-treated mice (Supplementary Fig. 1d). These data suggest that AHR controls Foxp3 expression.

We studied AHR expression by real-time PCR (RT-PCR) in T<sub>reg</sub> cells isolated from Foxp3<sup>gpf</sup> knockin mice. Foxp3<sup>gpf</sup> knockin mice have a GFP reporter inserted in the Foxp3 gene that facilitates the tracking and isolation of GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells<sup>18</sup>. Purified

<sup>1</sup>Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. <sup>2</sup>EMBL Outstation – Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. <sup>3</sup>Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, Massachusetts 02139, USA.

CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells showed increased expression levels of AHR, in agreement with reports of AHR expression by CD4<sup>+</sup>CD25<sup>+</sup> T cells<sup>19</sup> and AHR being part of the T<sub>reg</sub> cell signature<sup>20</sup>. We also detected the expression of the AHR-responsive gene *Cyp1a1* in CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells; treatment with the AHR antagonist resveratrol resulted in a significant decrease in CYP1A1 and Foxp3 expression ( $P < 0.0235$  and  $P < 0.0023$ , respectively, Fig. 1c) and suppressive activity ( $P < 0.05$ , Supplementary Fig. 2a), suggesting that the detected AHR is functional.

We then studied the effect of AHR activation by TCDD on T<sub>reg</sub> cells *in vivo*. Naive C57BL/6 mice were treated with TCDD (1 µg per mouse, intraperitoneal delivery) and immunized 24 h later with the 35–55 peptide from the myelin oligodendrocyte glycoprotein (MOG<sub>35–55</sub>) in complete Freund's adjuvant (CFA). Draining lymph nodes were prepared 10 days later and CD4<sup>+</sup>Foxp3<sup>+</sup> T<sub>reg</sub> cells were quantified by fluorescence-activated cell sorting (FACS). A single administration of TCDD followed by immunization with MOG<sub>35–55</sub> led to a significant ( $P < 0.001$ ) expansion of



**Figure 1 | AHR controls Foxp3 expression and T<sub>reg</sub> cell generation.**

**a**, Activation of the transcription of Renilla-luciferase-tagged Foxp3 (BACFoxp3:Ren) by expression in EL-4 cells of mouse AHR or a constitutively activated (CA) TGF-β receptor II. Renilla activity was normalized to the luciferase activity of a co-transfected control (mean + s.d. of triplicates). A construct coding for Renilla luciferase under the control of the herpes simplex virus thymidine kinase promoter (TK) was used as a positive control. **b**, ChIP analysis of the interaction of AHR with NCABS and CABS in Foxp3 and Cyp1a1 in CD4<sup>+</sup> T cells treated with TCDD (mean + s.d.). **c**, Ahr, Cyp1a1 and Foxp3 expression measured by real-time PCR on CD4<sup>+</sup>GFP-Foxp3<sup>-</sup> T cells (GFP<sup>-</sup>), CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells (GFP<sup>+</sup>), and CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells treated with resveratrol for 5 h (GFP<sup>+</sup> + R) (mean + s.d. of triplicates normalized to Gapdh expression,

data expressed as delta threshold cycle, ΔC<sub>t</sub>). **d**, Frequency of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells in the TCDD- or control-treated mice (mean + s.d. of three mice). **e**, MOG<sub>35–55</sub>-specific suppressive activity of T<sub>reg</sub> cells purified from TCDD- or control-treated mice, assayed using CD4<sup>+</sup>GFP-Foxp3<sup>-</sup> 2D2 T cells. CD4<sup>+</sup>GFP-Foxp3<sup>-</sup> T cells were used as responders (T<sub>eff</sub>). Cell proliferation is indicated as counts per min (c.p.m.) + s.d. in triplicate wells. **f**, Conversion of CD4<sup>+</sup>GFP-Foxp3<sup>-</sup> T cells into CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells by stimulation with antibodies to CD3 and CD28 in the presence of TCDD or TGF-β1. **g**, Suppressing activity of natural T<sub>reg</sub> cells or T<sub>reg</sub> cells induced with TGF-β1 or TCDD. Cell proliferation is indicated as c.p.m. + s.d. in triplicate wells. **h**, Inhibition of T<sub>reg</sub> cell conversion after activation with antibodies to CD3 and CD28 and treatment with TCDD or TGF-β1 in the presence of resveratrol (mean + s.d.).



CD4<sup>+</sup>Foxp3<sup>+</sup> T cells (Fig. 1d). The CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T cells expanded *in vivo* by TCDD administration and MOG<sub>35–55</sub> immunization were functional and showed increased MOG<sub>35–55</sub>-specific suppressive activity (Fig. 1e).

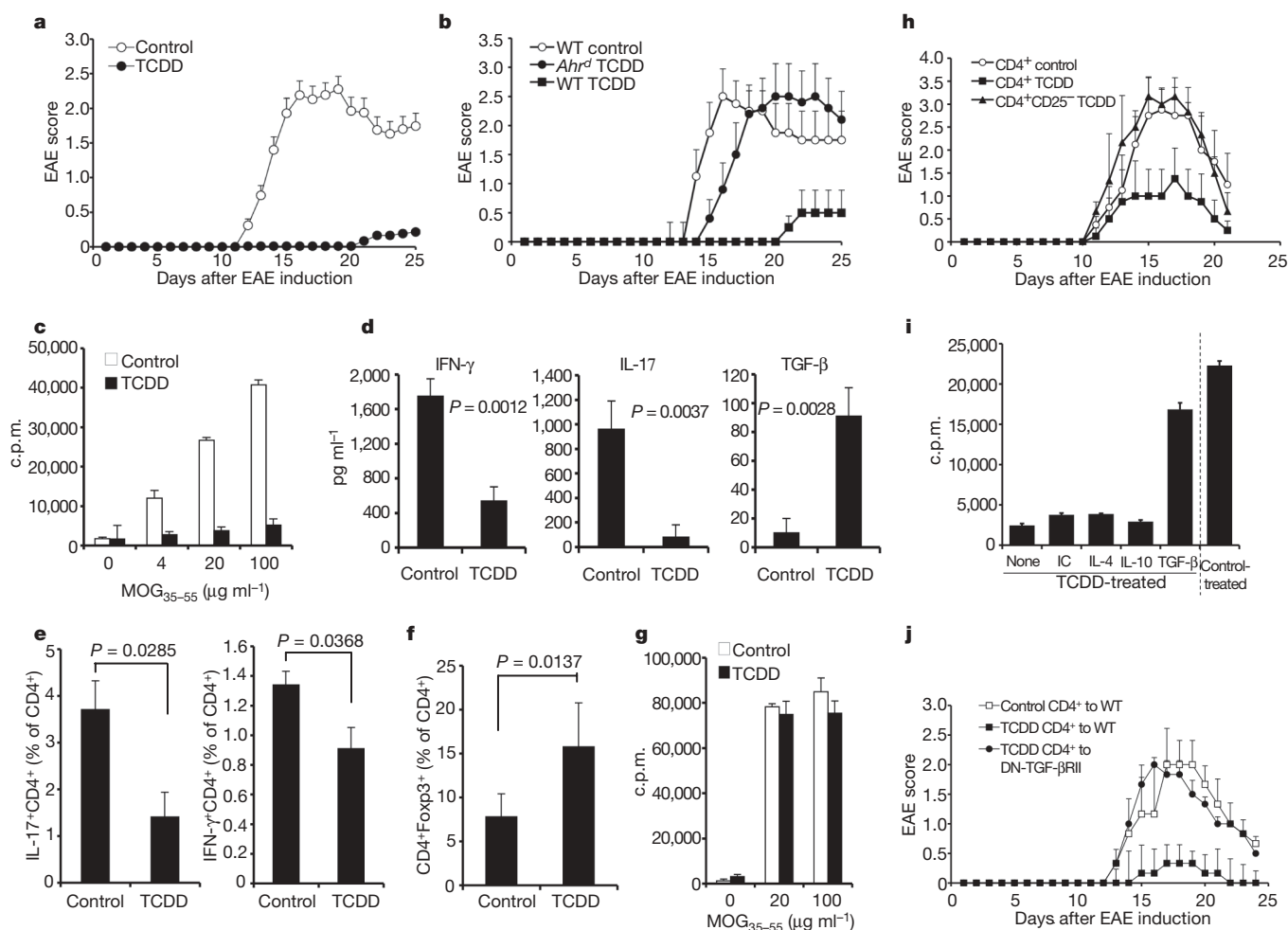
To establish if TCDD triggered the conversion of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells into new Foxp3<sup>+</sup> T<sub>reg</sub> cells, FACS-sorted CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T cells were activated *in vitro* with antibodies to CD3 and CD28 in the presence of TCDD, and the generation of CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells was followed by FACS. TGF-β1 was used as a positive control<sup>7</sup>. TCDD triggered the conversion of approximately 13% of the CD4<sup>+</sup> T cells in culture into CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells (Fig. 1f). CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells induced by TCDD showed a suppressive activity similar to that of T<sub>reg</sub> cells induced *in vitro* with TGF-β1 or CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells sorted from naive Foxp3<sup>gfp</sup> mice (Fig. 1g). Thus, AHR activation by TCDD can trigger the conversion of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells into functional CD4<sup>+</sup>GFP<sup>+</sup> T<sub>reg</sub> cells.

Treatment with the AHR antagonist resveratrol (50 μM) interfered with the induction of T<sub>reg</sub> cells by TGF-β1 and TCDD, but had a

stronger effect on the T<sub>reg</sub> cell conversion triggered by TCDD ( $P = 0.0053$ , Fig. 1h). CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells purified from naive mice did not proliferate on stimulation with antibodies to CD3 and CD28 and did not show increased suppressive activity after treatment with TCDD (Supplementary Fig. 2b, c), demonstrating that AHR is more important for the differentiation of new T<sub>reg</sub> cells than for the activity of established T<sub>reg</sub> cells.

### TCDD-induced T<sub>reg</sub> cells suppress EAE

To analyse the *in vivo* functionality of the T<sub>reg</sub> cells induced by AHR activation, we studied the effect of TCDD on EAE development. C57BL/6 mice were given a single intraperitoneal dose of TCDD and 1 day later EAE was induced with MOG<sub>35–55</sub>. TCDD had a dose-dependent effect on EAE induction; 1 μg per mouse markedly inhibited the clinical ( $P < 0.001$ , Fig. 2a, Supplementary Table 1 and Supplementary Fig. 3) and histopathological (Supplementary Fig. 4) signs of EAE. Oral administration of TCDD (1 μg per mouse) also prevented EAE development ( $P < 0.001$ ,  $n = 10$ ; data not shown). We confirmed that the effects of TCDD on EAE were mediated by



**Figure 2 | AHR activation by TCDD suppresses EAE by a TGF-β1-dependent mechanism.** **a**, EAE in TCDD- or control-treated mice (mean EAE score + s.e.m.). **b**, EAE in TCDD-treated C57BL/6 wild type (WT) or *Ahr*<sup>d</sup> mice (mean EAE score + s.e.m.). **c**, Recall proliferative response to MOG<sub>35–55</sub> of lymph node cells from TCDD- or control-treated mice (c.p.m. + s.d. in triplicate wells). **d**, Cytokine secretion (pg ml<sup>-1</sup>) triggered by MOG<sub>35–55</sub> in lymph node cells from TCDD- or control-treated mice. **e**, Frequency of CD4<sup>+</sup>IL-17<sup>+</sup> and CD4<sup>+</sup>IFN-γ<sup>+</sup> T cells in TCDD- or control-treated mice (mean percentage of cytokine-positive cells within the CD4<sup>+</sup>Foxp3<sup>+</sup> T-cell population + s.d., five mice were included per group). **f**, Frequency of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells in the spleen of TCDD- or control-

treated mice 21 days after EAE induction (mean + s.d. of five mice). **g**, Recall proliferative response to MOG<sub>35–55</sub> of CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> lymph node cells from TCDD- or control-treated Foxp3<sup>gfp</sup> mice (c.p.m. + s.d. in triplicate wells). **h**, EAE in mice transferred with CD4<sup>+</sup> or CD4<sup>+</sup>CD25<sup>-</sup> T cells from TCDD- or control-treated mice (EAE score + s.e.m.). **i**, Proliferation of CD4<sup>+</sup> T cells from TCDD-treated animals activated *in vitro* with MOG<sub>35–55</sub> and blocking antibodies to IL-4, IL-10 and TGF-β (c.p.m. + s.d. in triplicate wells). **j**, EAE in naive wild-type (WT) or dominant-negative TGF-βRII mice transferred with CD4<sup>+</sup> T cells from TCDD- or control-treated mice (mean EAE score + s.e.m.).

AHR by using C57BL/6 mice carrying the *d* allele of the *Ahr* gene (*Ahr<sup>d</sup>* mice). This allele codes for an AHR molecule with a reduced affinity (10–100-fold) for its ligands due to mutations in its ligand-binding site<sup>21</sup>. TCDD administration (1 µg per mouse) did not increase CD4<sup>+</sup>Foxp3<sup>+</sup> T<sub>reg</sub> cell levels (data not shown) and did not prevent EAE development in *Ahr<sup>d</sup>* mice (Fig. 2b, Supplementary Table 2 and Supplementary Fig. 5). Thus, AHR activation by TCDD inhibits the development of EAE.

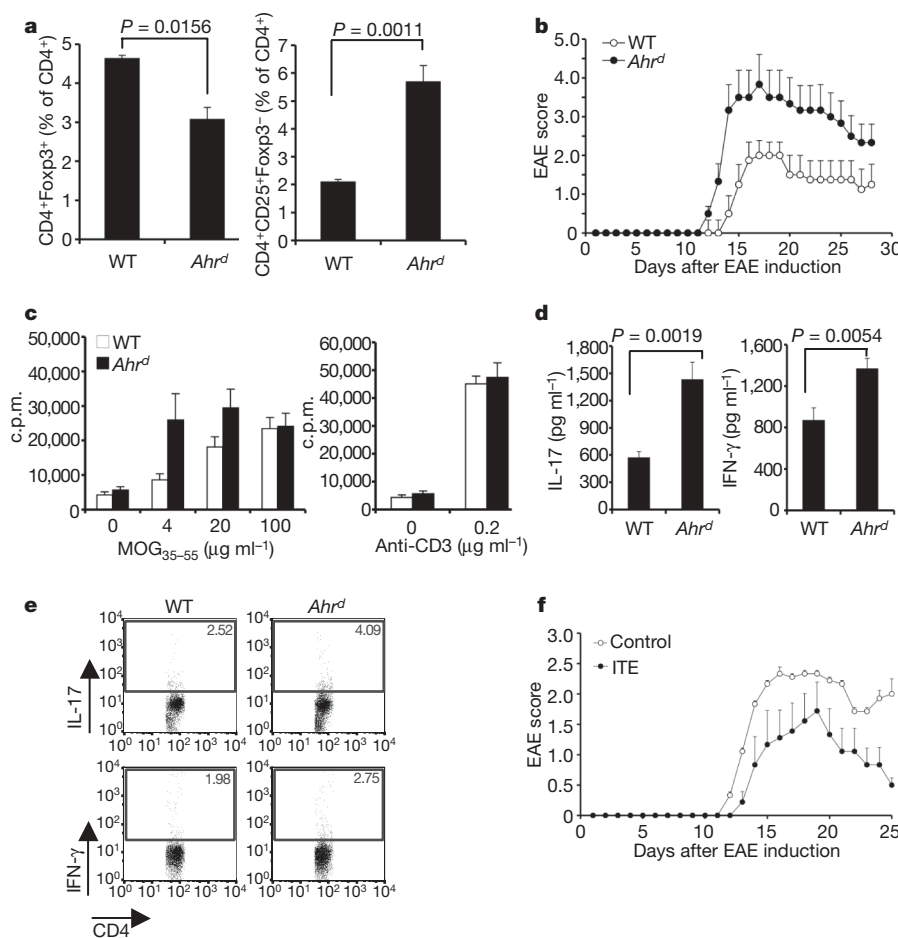
We analysed the immune response in TCDD-treated mice with antigen microarrays containing a collection of 362 CNS-related antigens (listed in Supplementary Table 3); similar arrays have recently identified new therapeutic targets in multiple sclerosis<sup>22</sup>. The control of EAE by TCDD correlated with a significant decrease in IgG reactivity to 97 antigens shown in Supplementary Fig. 6a and listed in Supplementary Table 4.

To characterize further the suppression of EAE by AHR activation with TCDD we studied the T-cell recall response in TCDD-treated mice. TCDD-treated mice showed a suppressed recall proliferative response to MOG<sub>35–55</sub> (Fig. 2c), but no differences were seen on activation with antibodies to CD3 (Supplementary Fig. 6b). When compared to the draining lymph node cells from control animals, cells from TCDD-treated mice secreted higher amounts of TGF-β1 and lower amounts of interferon (IFN)-γ and IL-17 on activation with MOG<sub>35–55</sub> (Fig. 2d); we did not detect significant amounts of IL-4 or IL-10 (data not shown). Moreover, AHR activation with TCDD led to a decrease in the frequency of CD4<sup>+</sup>IL-17<sup>+</sup> and CD4<sup>+</sup>IFN-γ<sup>+</sup> T cells in the draining lymph nodes (Fig. 2e).

The inhibition of EAE by AHR activation with TCDD was associated with a significant increase in the frequency of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells (Fig. 2f). Effector CD4<sup>+</sup>GFP-Foxp3<sup>–</sup> T cells purified from TCDD-treated mice showed normal proliferation (Fig. 2g) but significantly decreased secretion of IL-17 and IFN-γ on activation with MOG<sub>35–55</sub> (Supplementary Fig. 6c).

To confirm that the protective effect of TCDD on EAE was mediated by T<sub>reg</sub> cells, we depleted the natural T<sub>reg</sub> cells with antibodies to CD25 before TCDD treatment (Supplementary Fig. 7a). TCDD-treated mice showed a faster rebound in their T<sub>reg</sub> cell numbers ( $P < 0.04$  at day 7; Supplementary Fig. 7a), concomitant with a significant delay in the onset of EAE ( $P < 0.03$ ) and a significant reduction in IL-17<sup>+</sup>CD4<sup>+</sup> T cells in the draining lymph nodes ( $P < 0.03$ , Supplementary Fig. 7b, c). Moreover, the transfer of  $5 \times 10^6$  CD4<sup>+</sup> T cells from TCDD-treated mice significantly inhibited the development of EAE ( $P < 0.001$ , Fig. 2h, Supplementary Table 5 and Supplementary Fig. 8); the protective effect was lost when CD4<sup>+</sup>CD25<sup>+</sup> T cells were depleted ( $P < 0.001$ , Fig. 2h, Supplementary Table 5 and Supplementary Fig. 8). All in all, these data suggest that AHR activation by TCDD results in the generation of CD4<sup>+</sup>Foxp3<sup>+</sup> T<sub>reg</sub> cells that control the encephalitogenic response.

TGF-β1 has been linked to the suppressive activity of T<sub>reg</sub> cells *in vitro* and *in vivo*<sup>23</sup>. To evaluate the involvement of TGF-β1 in the inhibition of the recall response to MOG<sub>35–55</sub> by T<sub>reg</sub> cells (Fig. 2c), we activated lymph node cells from TCDD-treated mice in the presence of blocking antibodies to IL-4, IL-10 or TGF-β1. Blocking antibodies to TGF-β1, but not to IL-4 or IL-10, recovered



**Figure 3 | Endogenous AHR ligands control EAE development.** **a**, Frequency of CD4<sup>+</sup>Foxp3<sup>+</sup> T<sub>reg</sub> cells and CD4<sup>+</sup>CD25<sup>+</sup>Foxp3<sup>–</sup> T cells in the blood of wild-type C57BL/6 and *Ahr<sup>d</sup>* mice (mean + s.d.). **b**, EAE in wild-type C57BL/6 and *Ahr<sup>d</sup>* mice (mean EAE score + s.e.m.). **c**, Proliferation of lymph node cells from *Ahr<sup>d</sup>* and wild-type mice (c.p.m. + s.d. in triplicate wells).

**d**, Cytokine secretion (pg ml<sup>–1</sup>) triggered by MOG<sub>35–55</sub> in lymph node cells from *Ahr<sup>d</sup>* or wild-type mice. **e**, Frequency of Foxp3<sup>–</sup>, IL-17<sup>+</sup> and IFN-γ<sup>+</sup> positive CD4<sup>+</sup> T cells in MOG<sub>35–55</sub>-immunized *Ahr<sup>d</sup>* or wild-type mice. **f**, EAE in ITE- or control-treated mice (mean EAE score + s.e.m.).



the recall response to MOG<sub>35–55</sub> (Fig. 2i). To analyse the role of TGF- $\beta$ 1 in the control of EAE, we transferred CD4<sup>+</sup> T cells from TCDD-treated mice into mice expressing a T-cell-restricted dominant-negative variant of the TGF- $\beta$  receptor II—T cells from these mice are unresponsive to TGF- $\beta$ 1 (ref. 24). T<sub>reg</sub> cells from TCDD-treated mice protected wild-type mice from EAE, but were ineffective in mice harbouring T cells unresponsive to TGF- $\beta$ 1 ( $P < 0.001$ , Fig. 2j, Supplementary Table 6 and Supplementary Fig. 9). Thus, T<sub>reg</sub> cells induced by the activation of AHR with TCDD control EAE by a TGF- $\beta$ 1-dependent mechanism.

### Endogenous AHR ligands control T<sub>reg</sub> cell development *in vivo*

Our results on the control of T<sub>reg</sub> cell development by AHR activation suggest that endogenous AHR ligands<sup>17</sup> might participate in immune regulation. Indeed, we detected lower levels of CD4<sup>+</sup>Foxp3<sup>+</sup> T cells ( $P = 0.0156$ , Fig. 3a) and higher levels of CD4<sup>+</sup>CD25<sup>+</sup>Foxp3<sup>−</sup> T cells ( $P < 0.0011$ , Fig. 3a) in naive *Ahr*<sup>d</sup> mice. On immunization, *Ahr*<sup>d</sup> mice developed increased EAE severity ( $P < 0.0128$ , Fig. 3b, Supplementary Table 7 and Supplementary Fig. 10). In addition, *Ahr*<sup>d</sup> mice displayed an increased proliferative recall response to MOG<sub>35–55</sub> (Fig. 3c), concomitant with the secretion of higher levels of the pro-inflammatory cytokines IL-17 and IFN- $\gamma$  (Fig. 3d), and increased frequencies of IL-17<sup>+</sup> and IFN- $\gamma$ <sup>+</sup>CD4<sup>+</sup> T cells (Fig. 3e).

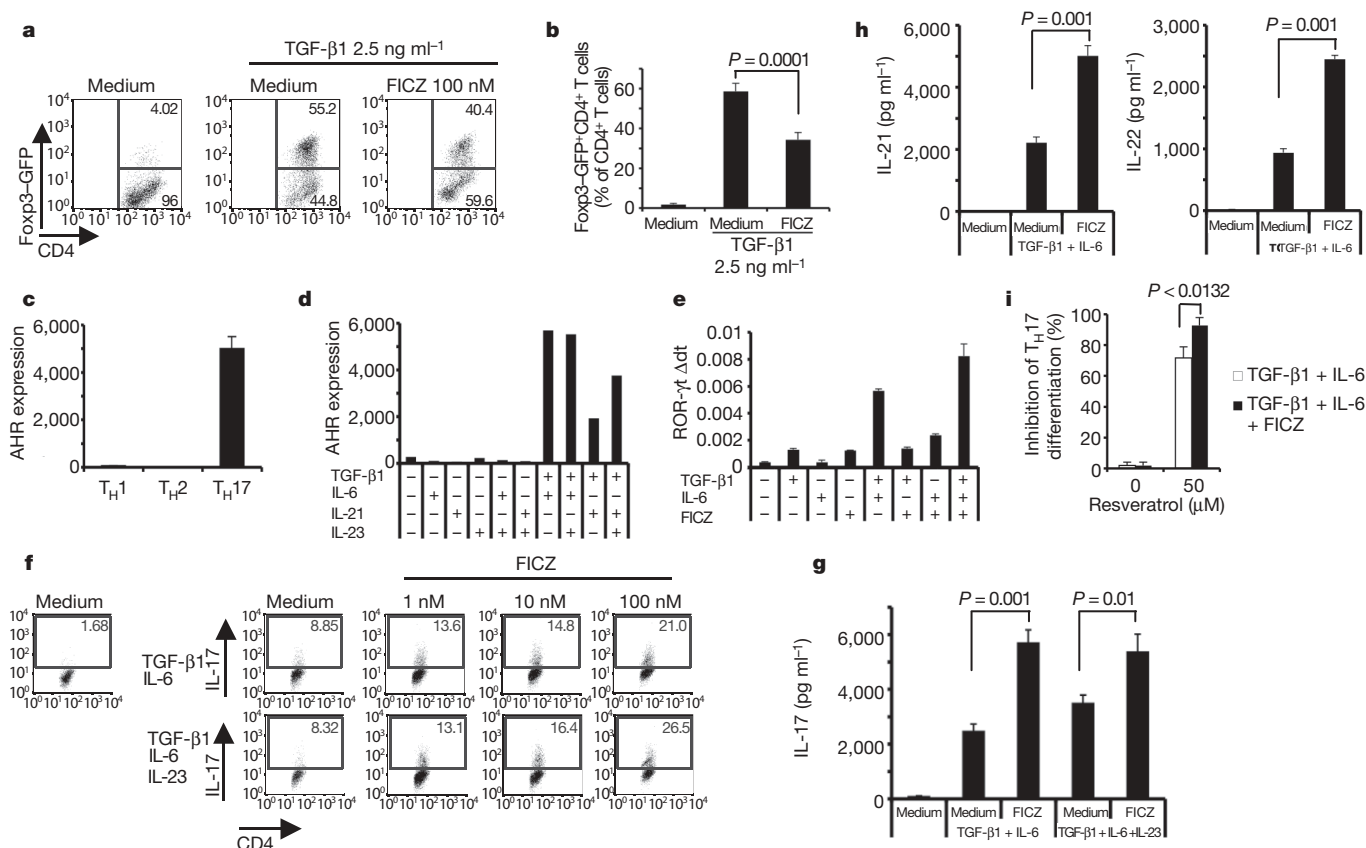
2-(1'H-indole-3'-carbonyl)-thiazole-4-carboxylic acid methyl ester (ITE)<sup>25</sup> is a tryptophan-derived AHR endogenous ligand. To study the physiological relevance of AHR activation for the control of T<sub>reg</sub> cell development, we analysed the effect of ITE on EAE. ITE

administration resulted in a significant reduction of EAE severity ( $P < 0.0137$ , Fig. 3f, Supplementary Table 8 and Supplementary Fig. 11), supporting a role for AHR endogenous ligands in immune regulation.

### Role of AHR on T<sub>H</sub>17 cell differentiation

It has been reported recently that keratinocytes exposed to ultraviolet-B (UV-B) light upregulate surface RANKL levels, signalling epidermal dendritic cells to support T<sub>reg</sub> cell expansion<sup>26</sup>. UV-B catalyses the formation of the AHR ligand FICZ *in vitro*, and is thought to generate FICZ in the skin<sup>27</sup>. Considering our findings on the role of AHR on T<sub>reg</sub> cells (Figs 1 and 2), and the reported effects of UV-B on T<sub>reg</sub> cell expansion<sup>26</sup>, we investigated the effect of FICZ on T<sub>reg</sub> cells. FICZ did not promote the conversion of CD4<sup>+</sup>GFP-Foxp3<sup>−</sup> T cells into CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells (data not shown); it actually interfered with the differentiation of T<sub>reg</sub> cells triggered by TGF- $\beta$ 1 *in vitro* ( $P < 0.0001$ , Fig. 4a, b). On the basis of the reciprocal relationship that exists between T<sub>H</sub>17 and T<sub>reg</sub> cells<sup>18</sup>, and the inhibitory effects of FICZ on T<sub>reg</sub> cell differentiation (Fig. 4a, b), we investigated the role of FICZ and AHR on T<sub>H</sub>17 cell differentiation.

We found that AHR expression is highly upregulated in T<sub>H</sub>17 T cells induced *in vitro* by activation with TGF- $\beta$ 1 and IL-6 (Fig. 4c). Moreover, AHR expression was also upregulated when T<sub>H</sub>17 cell differentiation was driven by TGF- $\beta$ 1, IL-6 and IL-23, or when IL-21 was used instead of IL-6 (ref. 9) (Fig. 4d). We then studied the effect of AHR activation by FICZ on *in vitro* T<sub>H</sub>17 cell differentiation. FICZ alone did not upregulate significantly the expression levels of



**Figure 4 | FICZ promotes T<sub>H</sub>17 cell differentiation.** **a, b**, Conversion of CD4<sup>+</sup>GFP-Foxp3<sup>−</sup> T cells into CD4<sup>+</sup>GFP-Foxp3<sup>+</sup> T<sub>reg</sub> cells by stimulation with antibodies to CD3 and CD28 in the presence of TCDD or by stimulation with TGF- $\beta$ 1 in the presence of FICZ. Error bars indicate mean  $\pm$  s.d. **c**, AHR expression in naive T cells differentiated *in vitro* into T<sub>H</sub>1, T<sub>H</sub>2 or T<sub>H</sub>17 cells for 4 days (mean  $\pm$  s.d. of triplicates normalized to actin expression). **d**, AHR expression in naive T cells differentiated *in vitro* into T<sub>H</sub>17 cells for 4 days with the indicated cytokines (mean  $\pm$  s.d. of triplicates

normalized to actin expression). **e**, ROR- $\gamma$ t expression in naive T cells differentiated into T<sub>H</sub>17 for 4 days with TGF- $\beta$ 1 and IL-6 alone or in combination with FICZ (mean  $\pm$  s.d. of triplicates normalized to actin expression). **f**, Frequency of IL-17<sup>+</sup> T cells differentiated with TGF- $\beta$ 1 and IL-6 alone or in combination with IL-23 and/or FICZ for 4 days. **g**, **h**, IL-17 (**g**) and IL-21 and IL-22 (**h**) levels in supernatants of cultures prepared as in **f** as measured by ELISA after 4 days of differentiation. **i**, Inhibition of T<sub>H</sub>17 cell differentiation by the AHR antagonist resveratrol (mean  $\pm$  s.d.).

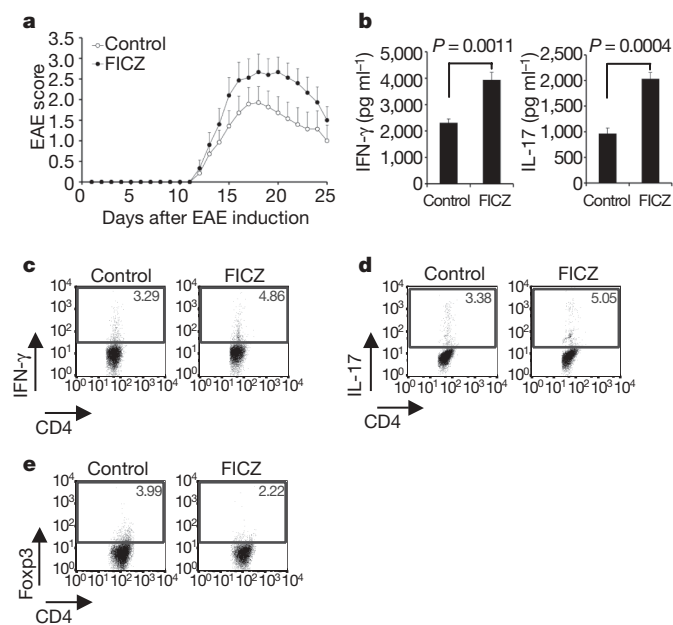
the  $T_H17$  transcription factor ROR- $\gamma$ t (Fig. 4e), but FICZ synergized with TGF- $\beta$ 1, IL-6 and IL-23 to drive  $T_H17$  cell differentiation *in vitro* (Fig. 4f–h). This synergism could be blocked with the AHR antagonist resveratrol (Fig. 4i).

We then tested the effect of FICZ *in vivo* in the EAE model. FICZ administration resulted in a significant worsening of EAE ( $P < 0.0038$ , Fig. 5a, Supplementary Table 9 and Supplementary Fig. 12), and was associated with increased frequencies of IL-17<sup>+</sup>CD4<sup>+</sup> and CD4<sup>+</sup>IFN- $\gamma$ <sup>+</sup> T cells (Fig. 5b–d), and increased secretion of IL-17 ( $P = 0.0004$ ) and IFN- $\gamma$  ( $P = 0.0011$ ) after *in vitro* stimulation with MOG<sub>35–55</sub> (Fig. 5b). Consistent with the reciprocal relationship that exists between  $T_H17$  and  $T_{reg}$  cells, FICZ-treated mice showed a decrease in the frequency of CD4<sup>+</sup>Foxp3<sup>+</sup>  $T_{reg}$  cells (Fig. 5e). Taken together, these results demonstrate that FICZ promotes  $T_H17$  cell differentiation *in vitro* and *in vivo*.

## Discussion

We have identified the dioxin receptor AHR as a regulator of  $T_{reg}$  and  $T_H17$  cell differentiation in a ligand-specific manner. AHR activation by TCDD has broad effects on the immune system<sup>28</sup>, including the induction of a T-cell population reminiscent of  $T_{reg}$  cells<sup>29</sup>.  $T_{reg}$  and  $T_H17$  cells are reciprocally related cell populations with opposite roles in the immune response<sup>18</sup>. TGF- $\beta$ 1 resembles the AHR in that it is a critical differentiation factor for both  $T_{reg}$  and  $T_H17$  cells; indeed AHR and TGF- $\beta$ 1 signalling pathways are known to cross-regulate each other in a cell-specific manner<sup>30,31</sup>. Thus, it is possible that AHR and TGF- $\beta$ 1 are part of a signalling pathway that drives a common cell precursor towards the  $T_{reg}$  or the  $T_H17$  cell differentiation programme. Moreover, because AHR activation resulted in the generation of TGF- $\beta$ 1-secreting cells, the effects of AHR on  $T_{reg}$  and  $T_H17$  cell differentiation might also involve the autocrine activity of TGF- $\beta$ 1.

Our findings demonstrate that AHR has opposite roles on  $T_H17$  and  $T_{reg}$  cell differentiation in a ligand-dependent fashion. AHR is known to associate and regulate the activity of other transcription factors like the retinoic acid<sup>32</sup> and the oestrogen receptor<sup>33</sup>, two receptors that influence  $T_{reg}$  and  $T_H17$  cell differentiation<sup>34,35</sup>.



**Figure 5 | FICZ worsens EAE.** **a**, EAE in FICZ- or control-treated mice (mean EAE score  $\pm$  s.e.m.). **b**, Cytokine secretion (pg ml<sup>-1</sup>) triggered by MOG<sub>35–55</sub> in splenocytes taken from FICZ or control-treated mice. **c–e**, Frequency of IFN- $\gamma$  (**c**), IL-17 (**d**) or Foxp3 (**e**) in splenocytes from FICZ- or control-treated mice.

Indeed, ligand-, cell-type- and activation-pathway-specific interactions of AHR with different transcriptional partners are known<sup>36–38</sup>. Thus, the promotion by AHR of both  $T_{reg}$  and  $T_H17$  cell differentiation might stem from its capacity to interact with different transcriptional partners in different cellular contexts. More importantly, the ability of AHR to regulate  $T_{reg}$  and  $T_H17$  cell differentiation in a ligand-specific manner makes it a unique target for the therapeutic manipulation of the immune response.

## METHODS SUMMARY

**Mice and reagents.** *Foxp3*<sup>gfp</sup> knockin mice and *Foxp3*<sup>gfp</sup> knockin 2D2 T-cell receptor (TCR) transgenic mice have been described<sup>39</sup>. C57BL/6, *Ahr*<sup>d</sup> and dominant-negative TGF- $\beta$ RII mice were purchased from Jackson Laboratories. TCDD and corn oil were purchased from Sigma-Aldrich. Antibodies to IL-10 and IL-4 were obtained from eBioscience and antibodies to TGF- $\beta$ 1 were obtained from R&D Systems.

**EAE induction.** EAE was induced by injecting the mice subcutaneously with 100  $\mu$ l of the MOG<sub>35–55</sub> peptide (MEVGWYRSPFSRVVHLYRNGK) in complete Freund's adjuvant oil. In addition, the mice received 150 ng of pertussis toxin (Sigma-Aldrich) intraperitoneally on days 0 and 2. Clinical signs of EAE were assessed according to the following score: 0, no signs of disease; 1, loss of tone in the tail; 2, hindlimb paresis; 3, hindlimb paralysis; 4, tetraplegia; 5, moribund.

**T-cell differentiation *in vitro*.** Naive CD4<sup>+</sup> T cells were stimulated for 4/5 days with plate-bound antibody to CD3 (145-2C11, 4  $\mu$ g ml<sup>-1</sup>) plus soluble antibody against CD28 (PV-1, 2  $\mu$ g ml<sup>-1</sup>) and recombinant cytokines: human TGF- $\beta$ 1 (2.5 ng ml<sup>-1</sup>, unless otherwise indicated), mouse IL-6 (30 ng ml<sup>-1</sup>). Polarization of T cells into  $T_H1$ ,  $T_H2$  or  $T_H17$  cells was performed with plate-bound antibody to CD3 (4  $\mu$ g ml<sup>-1</sup>) and soluble antibody to CD28 (2  $\mu$ g ml<sup>-1</sup>) in the presence of recombinant mouse IL-12 (10 ng ml<sup>-1</sup>; R&D Systems) plus anti-IL-4 (11B.11; 10  $\mu$ g ml<sup>-1</sup>) for  $T_H1$ , mouse IL-4 (10 ng ml<sup>-1</sup>; R&D Systems) plus anti-IL-12 (C17.8; 10  $\mu$ g ml<sup>-1</sup>) for  $T_H2$ , and TGF- $\beta$  plus IL-6 for  $T_H17$ . Monoclonal antibodies against mouse CD3, mouse CD28, mouse IL-4 and mouse IL-12 were purified from the supernatants of hybridomas obtained from the American Type Culture Collection (ATCC).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 October 2007; accepted 4 February 2008.

Published online 23 March 2008.

- Sakaguchi, S. Naturally arising CD4<sup>+</sup> regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.* **22**, 531–562 (2004).
- Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
- Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
- Brunkow, M. E. et al. Disruption of a new forkhead/winged-helix protein, scurfy, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nature Genet.* **27**, 68–73 (2001).
- Chatila, T. A. et al. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic dysregulation syndrome. *J. Clin. Invest.* **106**, R75–R81 (2000).
- Gavin, M. A. et al. Single-cell analysis of normal and FOXP3-mutant human T cells: FOXP3 expression without regulatory T cell development. *Proc. Natl Acad. Sci. USA* **103**, 6659–6664 (2006).
- Chen, W. et al. Conversion of peripheral CD4<sup>+</sup>CD25<sup>–</sup> naive T cells to CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells by TGF- $\beta$  induction of transcription factor Foxp3. *J. Exp. Med.* **198**, 1875–1886 (2003).
- Veldhoen, M., Hocking, R. J., Atkins, C. J., Locksley, R. M. & Stockinger, B. TGF $\beta$  in the context of an inflammatory cytokine milieu supports *de novo* differentiation of IL-17-producing T cells. *Immunity* **24**, 179–189 (2006).
- Korn, T. et al. IL-21 initiates an alternative pathway to induce proinflammatory  $T_H17$  cells. *Nature* **448**, 484–487 (2007).
- Ivanov, I. I. et al. The orphan nuclear receptor ROR $\gamma$ t directs the differentiation program of proinflammatory IL-17<sup>+</sup> T helper cells. *Cell* **126**, 1121–1133 (2006).
- Betelli, E., Oukka, M. & Kuchroo, V. K.  $T_H17$  cells in the circle of immunity and autoimmunity. *Nature Immunol.* **8**, 345–350 (2007).
- Ovcharenko, I. et al. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res.* **15**, 184–194 (2005).
- Okino, S. T., Quattrocchi, L. C., Pookot, D., Iwahashi, M. & Dahiya, R. A dioxin-responsive enhancer 3' of the human *CYP1A2* gene. *Mol. Pharmacol.* **72**, 1457–1465 (2007).
- Burchill, M. A., Yang, J., Vogtenhuber, C., Blazar, B. R. & Farrar, M. A. IL-2 receptor  $\beta$ -dependent STAT5 activation is required for the development of Foxp3<sup>+</sup> regulatory T cells. *J. Immunol.* **178**, 280–290 (2007).



15. Kim, H. P. & Leonard, W. J. CREB/ATF-dependent T cell receptor-induced FoxP3 gene expression: a role for DNA methylation. *J. Exp. Med.* **204**, 1543–1551 (2007).
16. Mantel, P. Y. *et al.* Molecular mechanisms underlying FOXP3 induction in human T cells. *J. Immunol.* **176**, 3593–3602 (2006).
17. Denison, M. S. & Nagy, S. R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334 (2003).
18. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T<sub>H</sub>17 and regulatory T cells. *Nature* **441**, 235–238 (2006).
19. Frericks, M., Meissner, M. & Esser, C. Microarray analysis of the AHR system: tissue-specific flexibility in signal and target genes. *Toxicol. Appl. Pharmacol.* **220**, 320–332 (2007).
20. Hill, J. A. *et al.* Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature. *Immunity* **27**, 786–800 (2007).
21. Okey, A. B., Vella, L. M. & Harper, P. A. Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1-450 by 3-methylcholanthrene. *Mol. Pharmacol.* **35**, 823–830 (1989).
22. Ousman, S. S. *et al.* Protective and therapeutic role for  $\alpha$ B-crystallin in autoimmune demyelination. *Nature* **448**, 474–479 (2007).
23. Li, M. O., Wan, Y. Y., Sanjabi, S., Robertson, A. K. & Flavell, R. A. Transforming growth factor- $\beta$  regulation of immune responses. *Annu. Rev. Immunol.* **24**, 99–146 (2006).
24. Gorelik, L. & Flavell, R. A. Abrogation of TGF $\beta$  signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12**, 171–181 (2000).
25. Song, J. *et al.* A ligand for the aryl hydrocarbon receptor isolated from lung. *Proc. Natl Acad. Sci. USA* **99**, 14694–14699 (2002).
26. Loser, K. *et al.* Epidermal RANKL controls regulatory T-cell numbers via activation of dendritic cells. *Nature Med.* **12**, 1372–1379 (2006).
27. Fritsche, E. *et al.* Lightening up the UV response by identification of the arylhydrocarbon receptor as a cytoplasmic target for ultraviolet B radiation. *Proc. Natl Acad. Sci. USA* **104**, 8851–8856 (2007).
28. Baccarelli, A. *et al.* Immunologic effects of dioxin: new results from Seveso and comparison with other studies. *Environ. Health Perspect.* **110**, 1169–1173 (2002).
29. Funatake, C. J., Marshall, N. B., Stepan, L. B., Mourich, D. V. & Kerkvliet, N. I. Cutting edge: activation of the aryl hydrocarbon receptor by 2,3,7,8-tetrachlorodibenzo-p-dioxin generates a population of CD4<sup>+</sup>CD25<sup>+</sup> cells with characteristics of regulatory T cells. *J. Immunol.* **175**, 4184–4188 (2005).
30. Wolff, S. *et al.* Cell-specific regulation of human aryl hydrocarbon receptor expression by transforming growth factor- $\beta$ 1. *Mol. Pharmacol.* **59**, 716–724 (2001).
31. Chang, X. *et al.* Ligand-independent regulation of transforming growth factor  $\beta$ 1 expression and cell cycle progression by the aryl hydrocarbon receptor. *Mol. Cell. Biol.* **27**, 6127–6139 (2007).
32. Murphy, K. A., Villano, C. M., Dorn, R. & White, L. A. Interaction between the aryl hydrocarbon receptor and retinoic acid pathways increases matrix metalloproteinase-1 expression in keratinocytes. *J. Biol. Chem.* **279**, 25284–25293 (2004).
33. Ohtake, F. *et al.* Modulation of oestrogen receptor signalling by association with the activated dioxin receptor. *Nature* **423**, 545–550 (2003).
34. Mucida, D. *et al.* Reciprocal T<sub>H</sub>17 and regulatory T cell differentiation mediated by retinoic acid. *Science* **317**, 256–260 (2007).
35. Tai, P. *et al.* Induction of regulatory T cells by physiological level estrogen. *J. Cell. Physiol.* **214**, 456–464 (2008).
36. Hestermann, E. V. & Brown, M. Agonist and chemopreventative ligands induce differential transcriptional cofactor recruitment by aryl hydrocarbon receptor. *Mol. Cell. Biol.* **23**, 7920–7925 (2003).
37. Oesch-Bartlomowicz, B. *et al.* Aryl hydrocarbon receptor activation by cAMP vs. dioxin: divergent signaling pathways. *Proc. Natl Acad. Sci. USA* **102**, 9218–9223 (2005).
38. Zhang, S., Rowlands, C. & Safe, S. Ligand-dependent interactions of the Ah receptor with coactivators in a mammalian two-hybrid assay. *Toxicol. Appl. Pharmacol.* **227**, 196–206 (2008).
39. Korn, T. *et al.* Myelin-specific regulatory T cells accumulate in the CNS but fail to control autoimmune inflammation. *Nature Med.* **13**, 423–431 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. I. Zon for helpful discussions and for the critical review of this manuscript; D. Kozoriz and J. P. Schatzmann Peron for help with FACS analysis; and D. Sherr and I. R. Cohen for useful discussions. This work was supported by grants AI435801 and NS38037 from the National Institutes of Health to H.L.W., TA3014A1/1 from the National Multiple Sclerosis Society to E.B. and R01AI073542-01 from the National Institutes of Health and RG-3882-A-1 from the National Multiple Sclerosis Society to M.O. F.J.Q. is a recipient of a long-term fellowship from the Human Frontiers of Science Program Organization; T.K. was supported by the Deutsche Forschungsgemeinschaft (KO 2964/1-1).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.W. ([hweiner@rics.bwh.harvard.edu](mailto:hweiner@rics.bwh.harvard.edu)).

## METHODS

**Real-time PCR.** RNA was extracted from cells using RNeasy columns (Qiagen), complementary DNA was prepared as recommended (Bio-Rad Laboratories) and used as template for real-time PCR. The expression of *Foxp3* was quantified with specific primers and probes (Applied Biosystems) on the GeneAmp 5500 Sequence Detection System (Applied Biosystems). Expression was normalized to the expression of the housekeeping genes *Gapdh* or *actin*.

**Generation of BACFoxp3:Ren.** The RP23-267C15 BAC clone contains 200 kilobases of mouse genomic DNA, including the entire locus of the *Foxp3* gene, and was used to target a *Renilla* cDNA cassette right after the ATG start codon of the *Foxp3* gene by homologous recombination using the Red recombineering system contained in the *Escherichia coli* DY380 bacterial strain.

**ChIP.** Cells were treated for 90 min with TCDD, fixed with 1% formaldehyde for 15 min and quenched with 0.125 M glycine. Chromatin was isolated and sheared to an average length of 300–500 base pairs by sonication. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. AHR-bound DNA sequences were immuno-precipitated with an AHR-specific antibody (Biomol SA-210). This antibody has been previously reported to work in ChIP assays<sup>36</sup>; we further validated its utility for our ChIP assays using chromatin isolated from TCDD-treated hepatocytes (data not shown). Crosslinks were reversed by incubation overnight at 65 °C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Quantitative PCR reactions were carried out using the following primer pairs: *Cyp1a1* –845 forward 5'-AGGCTCTTCTCAGCAACTC-3' and *Cyp1a1* –845 reverse 5'-CTGGGGC-TACAAAGGGTGAT-3'; *Foxp3* (NCAB-1) –2269 forward 5'-AGTGCCCA-TTACCTGTTAG-3' and *Foxp3* (NCAB-1) –2269 reverse 5'-GGAGGTCTGC-ATGGATCTTAG-3'; *Foxp3* (NCAB-2) –1596 forward 5'-GCCTTGTCAGG-AAAAACCTCTG-3' and *Foxp3* (NCAB-2) –1596 reverse 5'-GTCCTCGATTT-GGCACAGAC-3'; *Foxp3* (NCAB-3) –800 forward 5'-CTTGCCCTTCTTG-TGATG-3' and *Foxp3* (NCAB-3) –800 reverse 5'-TTGTGCTGAGTGCCC-TGAC-3'; *Foxp3* (CAB) +13343 forward 5'-GCTTTGTGCGAGTGAGAG-3' and *Foxp3* (CAB) +13343 reverse 5'-AGGGATTGGAGCACTTGTTG-3'. As a control, we amplified the *Untr6* region in chromosome 6 located at chromosome 6 120,258,582–120,258,797 using the following primers: *Untr6* forward 5'-TCAGGCATGAACCAACATAC-3' and *Untr6* reverse 5'-AACATCCACAG-TCCAGTGA-3'. Experimental  $C_t$  values were converted to copy numbers detected by comparison with a DNA standard curve run on the same PCR plates. Copy number values were then normalized for primer efficiency by dividing by the values obtained using input DNA and the same primer pairs. Error bars represent standard deviations calculated from the triplicate determinations.

**Antigen microarrays.** The antigens listed in Supplementary Table 1 were spotted onto Epoxy slides (TeleChem) as described<sup>40</sup>. Antigens were spotted in replicates of 6, the microarrays were blocked for 1 h at 37 °C with 1% bovine serum albumin, and incubated for 2 h at 37 °C with a 1:100 dilution of the test serum in blocking buffer. The arrays were then washed and incubated for 45 min at 37 °C with goat anti-mouse IgG Cy3-conjugated detection antibodies (Jackson ImmunoResearch Labs). The arrays were scanned with a ScanArray 4000X scanner (GSI Luminomics). Antigen reactivity was defined by the mean intensity of binding to the replicates of that antigen on the microarray. Raw data were normalized and analysed using GeneSpring software (Silicon Genetics) with the non-parametric Wilcoxon–Mann–Whitney *U*-test, using the Benjamini and Hochberg method with a false discovery rate (FDR) of 0.05 to determine significance<sup>41</sup>. The samples were clustered using a pairwise average linkage algorithm based on Spearman's rank correlation as a distance measure<sup>41</sup>.

**Cell proliferation and cytokine production.** Cells were cultured in serum-free X-VIVO 20 media (BioWhittaker) for 72 h. During the last 16 h, cells were pulsed with 1 mCi of [<sup>3</sup>H]thymidine (PerkinElmer) followed by harvesting on glass fibre filters and analysis of incorporated [<sup>3</sup>H]thymidine in a beta-counter (1450 Microbeta, Trilux, PerkinElmer). Alternatively, culture supernatants were collected 48 h after activation and the cytokine concentration was determined by ELISA using antibodies for IFN- $\gamma$ , IL-17, IL-4 and IL-10 from BD Biosciences and antibodies to TGF- $\beta$  from R&D Systems. For suppression assays, MACS-purified CD4<sup>+</sup>CD25<sup>+</sup> or FACS-purified CD4<sup>+</sup>GFP–*Foxp3*<sup>+</sup> T cells ( $1\text{--}5 \times 10^4$  cells well<sup>–1</sup>) were stimulated with antibodies to CD3 and C57BL/6 irradiated spleen cells ( $0.3\text{--}1.5 \times 10^5$  cells well<sup>–1</sup>) for 3 days in the presence of different ratios of CD4<sup>+</sup>GFP–*Foxp3*<sup>+</sup> T<sub>reg</sub> cells.

**Intracellular cytokine staining.** Cells were stimulated in culture medium containing 100 mg ml<sup>–1</sup> MOG<sub>35–55</sub> for 2 days or with PMA (50 ng ml<sup>–1</sup>) (Sigma-Aldrich) and ionomycin (1 mg ml<sup>–1</sup>) (Calbiochem) for 4 h; Golgistop (BD Biosciences) was added to the culture during the last 4 h. After staining of surface markers, cells were fixed and permeabilized using Cytofix/Cytoperm and Perm/Wash buffer from BD Biosciences according to the manufacturer's instructions. All antibodies to cytokines (IFN- $\gamma$ , IL-17, IL-10) including the corresponding isotype controls were obtained from BD Biosciences. Cells were incubated (1:100) at 25 °C for 20 min and washed twice in Perm/Wash before analysis. Data were acquired on a FACSCalibur (BD Biosciences) and analysed with FlowJo software (Tree Star).

40. Quintana, F. J. et al. Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc. Natl Acad. Sci. USA* 101 (suppl. 2), 14615–14621 (2004).

41. Stekel, D. *Microarray Bioinformatics* (Cambridge Univ. Press, Cambridge, 2003).



## LETTERS

# The sculpting of Jupiter's gossamer rings by its shadow

Douglas P. Hamilton<sup>1</sup> & Harald Krüger<sup>2,3</sup>

Dust near Jupiter is produced when interplanetary impactors collide energetically with small inner moons, and is organized into a main ring, an inner halo, and two fainter and more distant gossamer rings<sup>1,2</sup>. Most of these structures are constrained by the orbits of the moons<sup>3</sup> Adrastea, Metis, Amalthea and Thebe, but a faint outward protrusion called the Thebe extension behaves differently and has eluded understanding. Here we report on dust impacts detected during the Galileo spacecraft's traversal of the outer ring region<sup>4</sup>: we find a gap in the rings interior to Thebe's orbit<sup>5</sup>, grains on highly inclined paths, and a strong excess of submicrometre-sized dust just inside Amalthea's orbit. We present detailed modelling that shows that the passage of ring particles through Jupiter's shadow creates the Thebe extension and fully accounts for these Galileo results. Dust grains alternately charge and discharge when traversing shadow boundaries, allowing the planet's powerful magnetic field to excite orbital eccentricities<sup>6</sup> and, when conditions are right, inclinations as well.

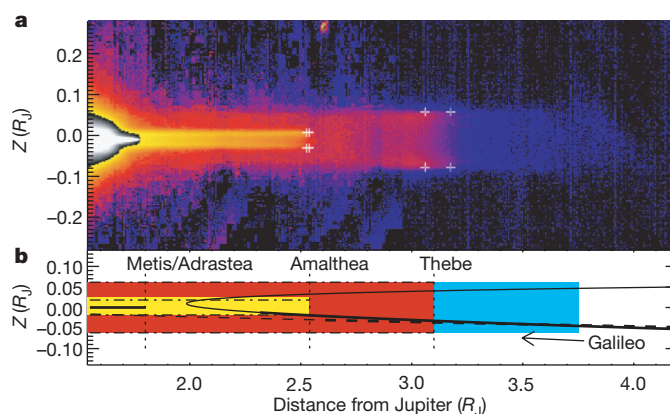
Jupiter's dusty rings were discovered by the Voyager spacecraft in 1979 and have subsequently been investigated with Earth- and space-based telescopes<sup>5,7</sup> as well as *in situ* spacecraft<sup>1,3,8,9</sup>. These observations have revealed a broad but very tenuous distribution of dust straddling the orbits of four small inner moons. For the most part, these moonlets bound the radial and vertical extent of the various ring structures, as summarized in Fig. 1a. There is, however, one major discrepancy: the Thebe extension, which has a vertical profile that matches Thebe's orbital excursions, but lies fully outside its source moonlet. Some dynamical process not considered by the standard ring models is clearly at work, but why does it single out the Thebe extension?

A unique source of data on the outer jovian rings was provided by Galileo, which passed through the tenuous Thebe and Amalthea rings in 2002 and 2003 (Fig. 1b). The spacecraft's impact-ionization dust detector<sup>10</sup> recorded ~100 impacts of micrometre- and submicrometre-sized dust grains, providing *in situ* information on particle sizes, speeds, impact directions and impact rates. Although the initial analysis of these data appeared in refs 4 and 11, it has not yet been reported in a scientific journal. The dust detector found a sharp transition between the Amalthea and Thebe rings in agreement with imaging, but other results were less expected. For example, some dust grains follow orbits tilted by ~20° from Jupiter's equator, which is surprising because particles in the visible rings have orbits tilted by only ~1°. The instrument also measured a reduction of particle flux interior to Thebe's orbit that was unanticipated because of the seemingly smooth increase of brightness of the optical image (red region of Fig. 1a; note that the vertical blue stripe is probably an artifact); see also ref. 5. Finally, the smallest detected dust grains, with radii of 0.2–0.3 µm, are far less prevalent near Thebe than near Amalthea, a difference that highlights the importance of electromagnetic forces<sup>12</sup>. These three results, detailed in Fig. 2a–c, nicely complement images of the

Thebe extension. The combined constraints allow us to demonstrate convincingly that electromagnetic processes involving Jupiter's shadow<sup>6</sup> govern the planet's gossamer rings.

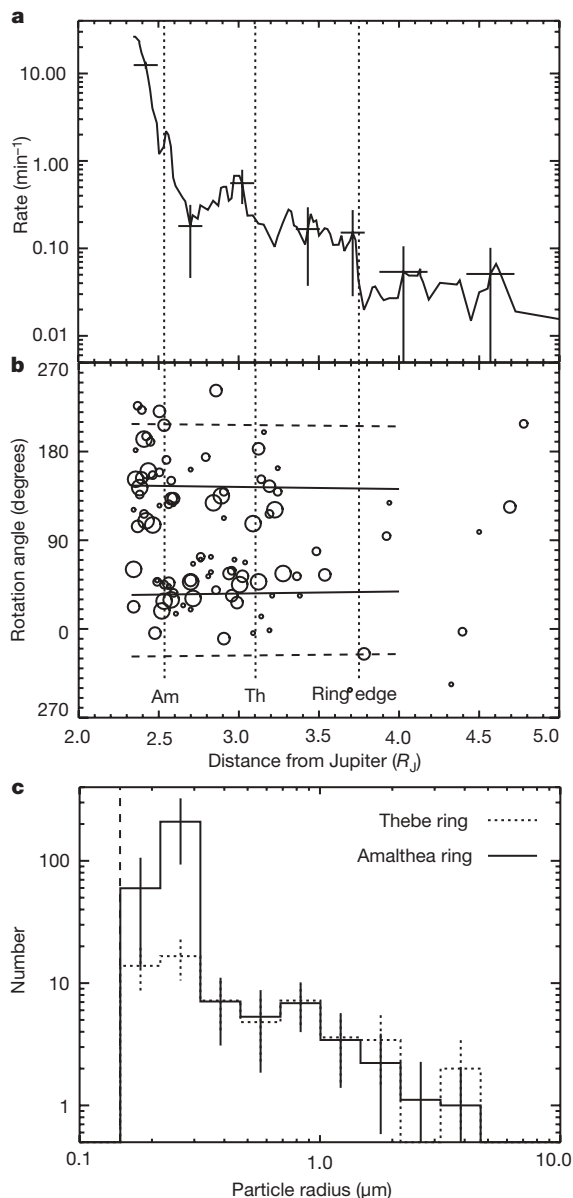
The orbits of micrometre-sized ring particles are dominated by Jupiter's gravity, but are also subject to electromagnetic forces and, to a far lesser extent, radiation pressure from the Sun. Dust grains acquire electric charges in Jupiter's vicinity by sweeping up local plasma and by photoemission<sup>6,13</sup> in which sunlight incident on a dust grain liberates electrons. Typically the charge is driven rapidly to an equilibrium value that depends on the dust-grain speed, the plasma composition and temperature, and the strength of solar radiation. These effects balance at one value in Jupiter's shadow and at another, larger, value in full sunlight.

Without planetary shadowing, the effects of electromagnetic forces on the sizes (semimajor axes  $a$ ) and shapes (eccentricities  $e$ ) of ring



**Figure 1 | Overview of Jupiter's diffuse ring system.** **a**, Jupiter's gossamer rings as seen by Galileo from a location near the planet's equatorial plane (false colour: adapted from ref. 1). Galileo observations<sup>1,3</sup> first showed that the ring system has at least four distinct components: (1) the main ring (in white) projecting inward from the two moonlets Metis and Adrastea, (2) the vertically extended halo (white) with thickness  $0.1R_J$  ( $R_J = 71,492$  km) interior to the main ring, (3) a gossamer ring associated with the satellite Amalthea (small yellow rectangle), and (4) an even fainter ring associated with Thebe (large red rectangle). Crosses mark the extremes of the radial and vertical motions of the two moons caused by their eccentric and inclined orbits. The thicknesses of the two outer rings exactly match the vertical excursions of the two source satellites, and ring material appears to extend primarily inward from all four moonlets<sup>1,3</sup>. The Thebe extension is the faint material (in blue) located outside Thebe's orbit. **b**, Galileo's trajectories during the ring passages on 5 November 2002 (solid line) and 21 September 2003 (dashed line). Dust data were acquired from most of the inbound legs (thick lines), but a spacecraft anomaly prevented measurements inside  $2.33R_J$  and during the outbound leg. We colour-code the ring components, and indicate the approximate locations of the moons' orbits by vertical dashed lines.

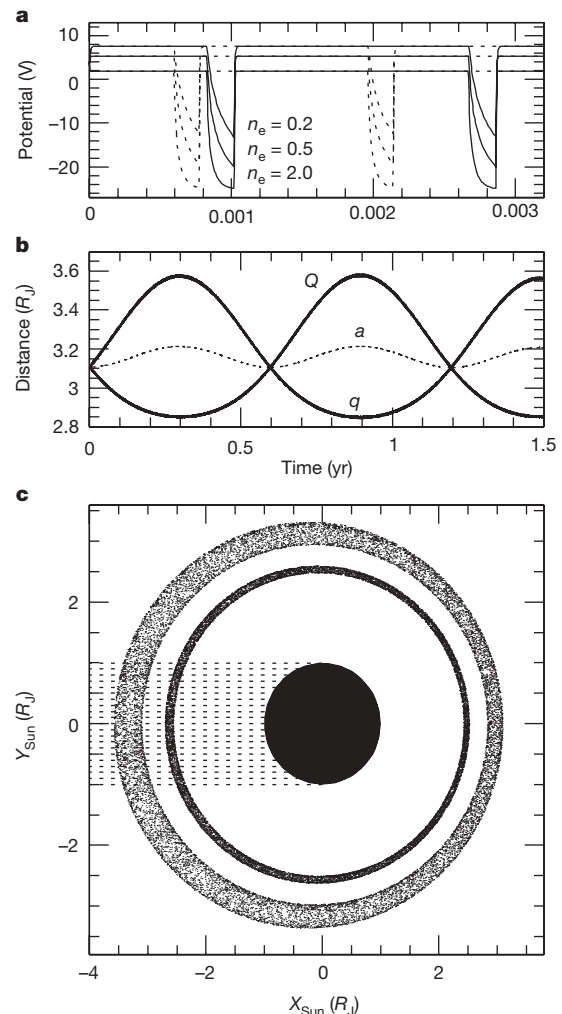
<sup>1</sup>Astronomy Department, University of Maryland, College Park, Maryland 20742-2421, USA. <sup>2</sup>Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany. <sup>3</sup>Max-Planck-Institut für Kernphysik, Postfach 103980, 69029 Heidelberg, Germany.



**Figure 2 | Galileo dust data.** **a**, Impact rate of submicrometre-sized dust particles onto the Galileo dust detector<sup>10</sup> during the 5 November 2002 ring passage. The rate increased by a factor of 1,000 as the spacecraft moved inward from 3.75  $R_J$  (right to left). The main rise at Amalthea's orbit (inner edge of the plot) is also seen in all other instrument channels, and corresponds to the transition from the Thebe to the Amalthea ring (Fig. 1). The next-most-significant feature is the dip in the rate by a factor of  $\sim 3$  from 3.0  $R_J$  to 2.7  $R_J$ . Although the error bars ( $\pm 1$  s.d.) are large, we believe that the feature is real, as it also appears in similar data for larger particles and in Galileo's second ring passage. Furthermore, the dip is also consistent with a detailed analysis of the Galileo images<sup>5</sup>. As indicated by the error bars, much of the fine structure at significantly lower impact rates is probably due to small number statistics. **b**, Distribution of dust grain trajectories. Larger symbols indicate larger particles. The rotation angle indicates the phase of the spinning spacecraft; the detector points closest to ecliptic north (south) at rotation angle 0° (180°). An analysis of the 140° field of view of the instrument combined with the 60° angle between the sensor and spin axes<sup>4,10,11</sup> and the spacecraft trajectory and spin axis pointing shows that particles on circular uninclined orbits could strike the detector when the spacecraft had a rotation angle between 30° and 150°. We indicate this detection limit with a pair of solid lines, and another for orbits with 20° tilts with dashed lines. A significant number of both large and small impacts occur outside the solid lines, indicating that some ring particles have large inclinations of  $\sim 20^\circ$ . **c**, Radii of detected dust grains in the Thebe ring (2.54–3.10  $R_J$ ; dotted) and the Amalthea ring (2.33–2.54  $R_J$ ; solid). Error bars,  $\pm 1$  s.d. The distributions are similar, except for a very strong excess of 0.2  $\mu\text{m}$  grains in the Amalthea ring.

particle orbits average to zero over satellite orbital periods ( $\sim 12$  h) (ref. 14). With shadowing, however, significant changes to semi-major axes and eccentricities can accumulate in phase over the  $\sim 6$  month timescale during which the long axes of elliptical orbits slowly rotate (or precess) in space. This effect is known as a shadow resonance<sup>6</sup>, and it produces a ring that is asymmetric and offset away from the Sun (Fig. 3c), similar to rings driven by radiation pressure<sup>15</sup>.

As the shadow resonance is a fully electromagnetic effect and magnetic fields do no work, energy measured in the reference frame that co-rotates with Jupiter's field is conserved. This constraint leads



**Figure 3 | Variations of the electric potential during shadow passages.**

**a**, Electric potential for 3.2  $\mu\text{m}$  dust grains from Amalthea (dotted lines) and Thebe (solid lines) as a function of time. We show charge histories for three different assumed plasma densities  $n_e$  in a simple two-component plasma model as dust grains travel from sunlight (flat curves) into Jupiter's shadow (dips to lower potentials) and out again. Dust in a plasma tends to charge negatively, and higher plasma densities lead to lower grain potentials, both in sunlight and in shadow. Sunlight raises equilibrium potentials because the photoelectric effect ejects electrons from a dust grain's surface. In reality, plasma properties probably vary with distance both from Jupiter and the equatorial plane, but we find that these non-uniform distributions lead to changes in orbital behaviour only over timescales much longer than those considered here, validating our simplifying assumption (compare ref. 19). **b**, The orbital evolution of a 3.2  $\mu\text{m}$  dust grain launched from Thebe into an  $n_e = 2.0 \text{ cm}^{-3}$  plasma; we plot the grain's pericentre distance  $q = a(1 - e)$ , semimajor axis  $a$ , and apocentre distance  $Q = a(1 + e)$ . Orbital evolution is due to the shadow resonance discussed in the text; note that dust spreads more easily outward from Thebe (at 3.1  $R_J$ ) than inward. **c**, A numerical simulation of the asymmetric Amalthea and Thebe rings formed by 3.2  $\mu\text{m}$  dust grains moving in an  $n_e = 2.0 \text{ cm}^{-3}$  plasma and crossing through the planet's shadow (stippled region).

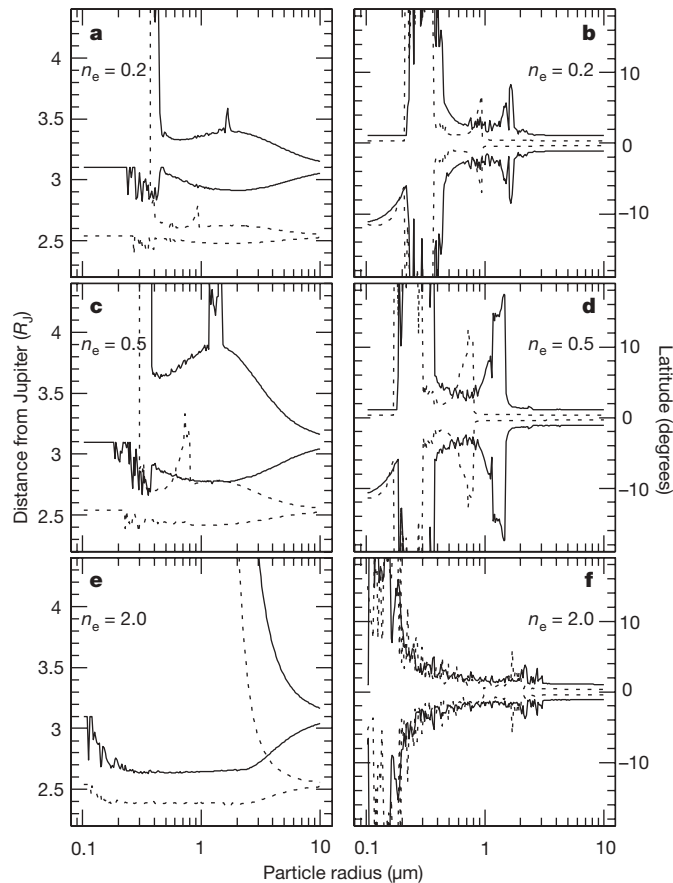


to a constant of the motion (the dimensionless Jacobi constant):

$$C_J = \frac{R_{\text{syn}}}{a} + 2\sqrt{\frac{a(1-e^2)}{R_{\text{syn}}}} \cos i$$

where  $i$  is a ring particle's orbital inclination and  $R_{\text{syn}} \approx 2.25R_J$  (here  $R_J$  is Jupiter's radius) is the radius of the synchronous orbit where a ring particle circles Jupiter at the same rate that the planet rotates<sup>6</sup>.

Particles blasted off satellites start on nearly circular orbits around Jupiter, because satellite orbital speeds ( $\sim 25 \text{ km s}^{-1}$ ) completely



**Figure 4 | Radial and vertical mobility of dust.** Here we plot numerically determined minimum and maximum distances (left plots) and latitudes (right plots) reached by different-sized ring particles within two years of launch from Thebe (solid curves) and Amalthea (dotted curves). Three different charging histories parameterized by  $n_e$  and 600 separate numerical simulations are presented here. The mobility of dust grains is enhanced by two factors: (1) small particle sizes which have larger electromagnetic perturbations, and (2) slow pericentre precession rates which allow the shadow resonance to pump eccentricities to extremely high values<sup>6</sup>. The total precession rate is the sum of a constant gravitational term and an electromagnetic contribution that is dependent on particle size and electric charge<sup>14</sup>. The two terms have the same sign for low plasma densities and are comparable in strength for particle radii of  $\sim 2 \mu\text{m}$  (near the peaks in the radial plots). For larger particles, electromagnetic perturbations weaken and radial mobility decreases. Smaller particles precess so rapidly (owing to electromagnetic perturbations) that changes to eccentricity can add together in phase for only short periods of time and again mobility decreases. **a, b,** For  $n_e = 0.2 \text{ cm}^{-3}$ , dust grains larger than  $\sim 0.5 \mu\text{m}$  cannot reach outward to the distance observed in Fig. 1 ( $3.75R_J$ ) nor can they attain latitudes in excess of  $10^\circ$ . **c, d,** With a plasma density of  $0.5 \text{ cm}^{-3}$ , micrometre-sized dust from Thebe can extend out to  $\sim 4R_J$  and attain latitudes of nearly  $20^\circ$ , as required by the observations. **e, f,** For  $n_e = 2.0 \text{ cm}^{-3}$ , micrometre-sized dust grains cannot reach large latitudes and dust grains smaller than  $2\text{--}3 \mu\text{m}$  escape from the ring entirely. An extremely slow pericentre precession rate, caused by nearly equal and opposite gravitational and electromagnetic terms, allows the shadow resonance to facilitate escape.

dominate the small ejection speeds ( $\sim 50 \text{ m s}^{-1}$ ). In the Thebe ring,  $a > R_{\text{syn}}$  and dynamical evolution leads to  $e \gg i$ , as the dominant  $\mathbf{v} \times \mathbf{B}$  forces are primarily radial (here  $\mathbf{v}$  is the velocity of the dust grain relative to Jupiter's magnetic field,  $\mathbf{B}$ ). Thus the equation shows that the semimajor axis and eccentricity increase in tandem, causing dust to spread more easily outward from Thebe than inward, as indicated by Fig. 3b. In fact, dust launched on a circular orbit from Thebe at  $a/R_{\text{syn}} = 1.38$  has  $C_J = 3.074$ ; by simultaneously taking  $a \rightarrow \infty$  and  $e = 1$ , we find that the smallest possible pericentre distance is  $q = a(1 - e) = 2.66R_J$ , about  $0.4R_J$  inside Thebe's orbit at  $3.1R_J$ , while the most distant orbit point  $Q = a(1 + e)$  can grow without bound. Thus, perturbations from an arbitrary time-varying dust grain charge can bring the dust grain at most  $0.4R_J$  inward from Thebe or about halfway to synchronous orbit. This hard limit on the inner extent of the Thebe dust ring provides a natural explanation for the existence of the gap in the dust number density measurements visible at  $\sim 2.7R_J$  in Fig. 2a.

We ran a number of dynamical simulations including gravitational, electromagnetic and radiation forces<sup>14</sup> to test our analytic prediction of a gap in the ring particle number density. A critical parameter in these simulations is the plasma density, which has not been measured. Accordingly, we ran the model for a number of different plasma densities ( $n_e$ ) and display three of them ( $n_e = 0.7, 0.5$  and  $2.0 \text{ cm}^{-3}$ ) in Fig. 4. All of the numerical simulations support the existence of an inner boundary, despite the fact that radiation pressure prevents the energy integral from being strictly obeyed. Over much longer timescales, diffusion induced by radiation pressure or dissipative Poynting-Robertson drag can cause material to leak inward from Thebe. Alternatively, ring material detected inside  $2.7R_J$  may simply have originated from Amalthea, as suggested by Fig. 4c.

We now compare this suite of models to the other observations discussed above. The visual appearance of the Thebe extension in Fig. 1a requires grains with radii larger than  $\sim 2 \mu\text{m}$  (ref. 5), significant radial extensions, and low orbital inclinations. The latitude plots of Fig. 4 indicate that particles of the right size do maintain low inclinations for all three plasma densities, but for the most tenuous  $n_e = 0.2 \text{ cm}^{-3}$  plasma no micrometre-sized material should reach the outer edge of the Thebe extension at  $3.75R_J$ . Requiring that the shadow resonance be strong enough to push material out to  $3.75R_J$  therefore puts a lower limit on the plasma number density, which in our simple model is  $n_e > 0.5 \text{ cm}^{-3}$ . Only the models shown in Fig. 4c–f are viable.

There are several reasons why the Amalthea ring does not show a similar extension<sup>5</sup>. First, the electromagnetic force (and hence the shadow resonance) is significantly weaker at Amalthea than at Thebe (note the less perturbed Amalthea ring in Fig. 3c), and in fact vanishes just  $\sim 0.3R_J$  inward at synchronous orbit where a ring particle's velocity relative to the magnetic field is zero. In addition, dust grains on crossing orbits survive against recollision with satellites for only  $\sim 100 \text{ yr}$  near Amalthea, but 25 times longer near Thebe owing to that satellite's smaller size, larger orbital inclination, and greater distance from Jupiter. Finally, an Amalthea extension less than a few per cent of the brightness of the Amalthea ring would be masked by the background debris from Thebe.

Galileo's dust detector observed a significant population of particles with orbits tilted by  $\geq 20^\circ$ . The models indicate that for certain plasma conditions, particle orbits do in fact attain large inclinations (Fig. 4b, d, f). The mechanism forcing the larger grains to high latitudes is a vertical shadow resonance, in which the line of nodes (the intersection of the orbit plane with Jupiter's equatorial plane) rotates extremely slowly, allowing inclinations to be pumped systematically. Similar to the standard shadow resonance, this requires a near cancellation between gravitational and electromagnetic perturbations to the nodal precession rate, which occurs for dust grains with a specific charge-to-mass ratio. Outside synchronous orbit, positive orbit-averaged charges<sup>14</sup> and therefore low plasma densities are needed.

The vertical shadow resonance forces micrometre and submicrometre grains onto orbits inclined by up to  $20^\circ$ , nicely matching the observed dust trajectories (Fig. 2b). Positive orbit-averaged charges indicate a plasma density  $n_e < 2.0 \text{ cm}^{-3}$  in our simple model. Only the models shown in Fig. 4c, d satisfy both this and the previous constraint.

Finally, Galileo observed a strong excess of  $0.2\text{--}0.3 \mu\text{m}$  dust grains at Amalthea (Fig. 2c). Figure 4 shows that, for all modelled plasma densities, electromagnetic forces cause positively-charged particles smaller than  $\sim 0.3 \mu\text{m}$  to escape entirely from the ring region<sup>12,16</sup>. For our favoured plasma density of  $n_e = 0.5 \text{ cm}^{-3}$  (Fig. 4c, d), the distance plot shows that a narrow size range from  $0.2$  to  $0.3 \mu\text{m}$  is retained in the Amalthea ring but escapes from the Thebe ring. This occurs because the shadow resonance, like all electromagnetic effects on ring particles, is weaker near the synchronous orbit just interior to Amalthea. Thus Thebe's extra loss mechanism explains the differences in the size distributions observed by Galileo.

We find that the shadow resonance at Jupiter can simultaneously account for (1) the outward extension of the Thebe ring at constant inclination, (2) the gap interior to Thebe, (3) the large number of high-inclination particles, and (4) the excess of  $0.2\text{--}0.3 \mu\text{m}$  grains at Amalthea in a natural model with only a single tunable parameter: the plasma density, which we constrain to  $0.5 \text{ cm}^{-3} < n_e < 2.0 \text{ cm}^{-3}$ . Changing other parameters of the plasma model will have similar effects to adjusting  $n_e$ , as ring particle dynamics is almost entirely determined by the details of the charging history (Fig. 3a). Shadow resonances are active to some extent in all diffuse dusty rings, and the vertical shadow resonance is a very promising mechanism for driving micrometre-sized ring particles to large latitudes. The origin of particles detected by the Voyager spacecraft at latitudes up to  $30^\circ$  near Saturn<sup>17,18</sup> remains an unsolved problem; we suggest that the vertical shadow resonance is a plausible solution.

Received 16 August 2007; accepted 28 February 2008.

1. Burns, J. A. *et al.* The formation of Jupiter's faint rings. *Science* **284**, 1146–1150 (1999).
2. Burns, J. A. *et al.* in *Jupiter: Planet, Satellites & Magnetosphere* (eds Bagenal, F., McKinnon, B. & Dowling, T.) 241–262 (Cambridge Univ. Press, Cambridge, UK, 2003).

3. Ockert-Bell, M. E. *et al.* The structure of Jupiter's ring system as revealed by the Galileo imaging experiment. *Icarus* **138**, 188–213 (1999).
4. Krüger, H. *Jupiter's Dust Disc, an Astrophysical Laboratory*. Thesis, Ruprecht-Karls-Universität (Shaker, Aachen, 2003).
5. Showalter, M. R., de Pater, I., Verbanac, G., Hamilton, D. P. & Burns, J. A. Properties and dynamics of Jupiter's gossamer rings from Galileo, Voyager, Hubble and Keck images. *Icarus*. (in the press).
6. Horányi, M. & Burns, J. A. Charged dust dynamics — Orbital resonance due to planetary shadows. *J. Geophys. Res.* **96**, 19283–19289 (1991).
7. de Pater, I. *et al.* Keck infrared observations of Jupiter's ring system near Earth's 1997 ring plane crossing. *Icarus* **138**, 214–223 (1999).
8. Throop, H. B. *et al.* The jovian rings: New results derived from Cassini, Galileo, Voyager and Earth-based observations. *Icarus* **172**, 59–77 (2004).
9. Showalter, M. R. *et al.* A new look at Jupiter's ring system: Preliminary results from New Horizons. *AAS/Div. Dyn. Astron. Meet.* **38**, no. 12.07 (2007).
10. Grün, E. *et al.* The Galileo dust detector. *Space Sci. Rev.* **60**, 317–340 (1992).
11. Moissl, R. *Galileo's Staubbmessungen in Jupiters Gossamer-Ringen*. Thesis, Ruprecht-Karls-Universität (2005).
12. Hamilton, D. P. & Burns, J. A. Ejection of dust from Jupiter's gossamer ring. *Nature* **364**, 695–699 (1993).
13. Schaffer, L. & Burns, J. A. Stochastic charging of dust grains in planetary rings: Diffusion rates and their effects on Lorentz resonances. *J. Geophys. Res.* **100**, 213–234 (1995).
14. Hamilton, D. P. Motion of dust in a planetary magnetosphere: Orbit-averaged equations for oblateness, electromagnetic, and radiation forces with application to Saturn's E ring. *Icarus* **101**, 244–264 (1993).
15. Hamilton, D. P. The asymmetric time-variable rings of Mars. *Icarus* **119**, 153–172 (1996).
16. Horányi, M., Morfill, G. E. & Grün, E. Mechanism for the acceleration and ejection of dust grains from Jupiter's magnetosphere. *Nature* **363**, 144–146 (1993).
17. Gurnett, D. A., Grün, E., Gallagher, D., Kurth, W. S. & Scarf, F. L. Micron-sized particles detected near Saturn by the Voyager plasma wave instrument. *Icarus* **53**, 236–254 (1983).
18. Srama, R. *et al.* In situ dust measurements in the inner Saturnian system. *Planet. Space Sci.* **54**, 967–987 (2006).
19. Horányi, M. & Cravens, T. E. The structure and dynamics of Jupiter's ring. *Nature* **381**, 293–298 (1996).

**Acknowledgements** We thank the Galileo project for a successful mission. This research was supported by grants from NASA Exobiology (D.P.H.) and DLR (H.K.).

**Author Contributions** H.K. did all of the data analysis and created Figs 1 and 2. D.P.H. did the analytic work and numerical modelling, including Figs 3 and 4.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.P.H. ([dphamil@umd.edu](mailto:dphamil@umd.edu)).



## LETTERS

# Phase-slip-induced dissipation in an atomic Bose–Hubbard system

D. McKay<sup>1</sup>, M. White<sup>1</sup>, M. Pasienski<sup>1</sup> & B. DeMarco<sup>1</sup>

Phase-slips control dissipation in many bosonic systems, determining the critical velocity of superfluid helium<sup>1</sup> and the generation of resistance in thin superconducting wires<sup>2</sup>. Technological interest has been largely motivated by applications involving nanoscale superconducting circuit elements, such as standards based on quantum phase-slip junctions<sup>3</sup>. Although phase slips caused by thermal fluctuations at high temperatures are well understood<sup>4</sup>, controversy remains over the role of phase slips in small-scale superconductors<sup>5</sup>—in solids, problems such as uncontrolled noise sources and disorder complicate their study and application<sup>6</sup>. Here we show that phase slips can lead to dissipation in a clean and well-characterized Bose–Hubbard system, by experimentally studying the transport of ultracold atoms trapped in an optical lattice. In contrast to previous work, we explore a low-velocity regime described by the three-dimensional Bose–Hubbard model that is unaffected by instabilities, and we measure the effect of temperature on the dissipation strength. The damping rate of atomic motion (the analogue of electrical resistance in a solid) in the confining parabolic potential is well fitted by a model that includes finite damping at zero temperature. The low-temperature behaviour is consistent with the theory of quantum tunnelling of phase slips, whereas at higher temperatures a crossover consistent with a transition to thermal activation of phase slips is evident. Motion-induced features reminiscent of vortices and vortex rings associated with phase slips are also observed in time-of-flight imaging. These results clarify the role of phase slips in superfluid systems. They may also be of relevance in understanding the source of metallic phases observed in thin films<sup>7,8</sup>, or serve as a test bed for theories of bosonic dissipation based upon variants of the Bose–Hubbard model<sup>9</sup>.

Although it is widely believed to be the simplest model that captures the relevant features of boson physics in a variety of physical systems, the Bose–Hubbard model is not integrable, and a full characterization of its features is therefore a challenging theoretical problem. The Bose–Hubbard model is described by the hamiltonian

$$H = -J \sum_{\langle ij \rangle} (b_i^\dagger b_j + b_i b_j^\dagger) + \sum_i \varepsilon_i n_i + \frac{U}{2} \sum_i n_i(n_i - 1) \quad (1)$$

where  $b_i$  operates to remove an atom from site  $i$ , the notation  $\langle ij \rangle$  indicates pairs of sites that are nearest neighbours,  $n_i = b_i^\dagger b_i$  is the number of bosons on site  $i$ , and  $\varepsilon_i$  is the energy cost for a boson to occupy site  $i$ . Particles in the Bose–Hubbard model move by tunnelling with energy  $J$  between adjacent lattice sites and interact pairwise on the same site with energy  $U$ . In our experiment we can directly test this model because ultracold atoms trapped in an optical lattice are a realization of the Bose–Hubbard model (for sufficient lattice potential depth)<sup>10,11</sup>. Although material and electronic parameters are not easy to configure independently in solid systems, we are able to quantitatively probe the Bose–Hubbard

model by precisely determining and controlling parameters such as  $J$  and  $U$ .

We study mass transport, which is the equivalent of charge transport for neutral bosons. Charge transport is studied in solids by using an electric field to apply a uniform force to the charge carriers. In contrast, we use the spatially inhomogeneous restoring force from a parabolic confining potential to excite damped harmonic motion of the centre of mass (COM) of the gas of atoms. The confining potential is included in equation (1) through the site energies  $\varepsilon_i = kr_i^2/2$ , where  $r_i$  is the distance from site  $i$  to the centre of the harmonic potential and  $k$  is the spring constant of the restoring force. We measure the COM motion damping rate for small COM velocities. In this regime, which is analogous to the linear, or ohmic, regime for conductivity experiments on solids, the damping rate is independent of velocity. In analogy to experiments on solids, we measure the response of the damping rate to changes in the temperature of the gas of atoms. We also examine how the damping rate depends on the ratio  $J/U$ , which is controlled by tuning the lattice potential depth. This dependence cannot easily be studied in solids.

Previous work on the transport of Bose–Einstein condensates (BECs) in optical lattices has focused on regimes not described by the Bose–Hubbard model<sup>12</sup>, on low-dimensional systems<sup>13,14</sup>, on relatively high-velocity transport<sup>15</sup> and on probing the Landau and dynamic instabilities<sup>16–19</sup>. In contrast to much of that research, we work at low velocity and do not study phenomena associated with either mentioned instability. The maximum COM velocities in our data are controlled to be smaller than the critical velocity for the Landau and dynamic instabilities<sup>20</sup>, including the effect of strong interactions<sup>21</sup>. We do not observe any phenomena characteristic of these instabilities, such as a significant change in the condensate fraction, strong nonlinear damping, or excitations of the condensate similar to those observed in ref. 17. Measurements of temperature-dependent dissipation in solids have proven to be a powerful tool in understanding phenomena such as phase slips<sup>5</sup> and other sources of resistance<sup>7</sup>; our work is the first systematic investigation of the effect of temperature on transport in an optical lattice.

The experimental sequence is shown in Fig. 1. A <sup>87</sup>Rb BEC is created in a Ioffe–Pritchard magnetic trap using standard cooling and trapping techniques (see ref. 22 for a description of our BEC apparatus). The BEC is adiabatically transferred into a cubic optical lattice with  $d = 405.9$  nm between sites; three atoms or fewer are confined on each lattice site for the data used in this paper. The strength of the lattice potential is characterized by a dimensionless parameter  $s$  defined by the lattice potential depth  $sE_R$  along each lattice direction ( $E_R = \hbar^2 \pi^2 / 2md^2$  is the recoil energy and  $m$  is the atomic mass). After the BEC is transferred into the lattice, COM motion is generated by applying a rapid impulse to the BEC along the vertical direction  $z$ . The COM velocity is measured using time-of-flight imaging after the motion has been allowed to freely evolve for up to 200 ms.

<sup>1</sup>Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA.

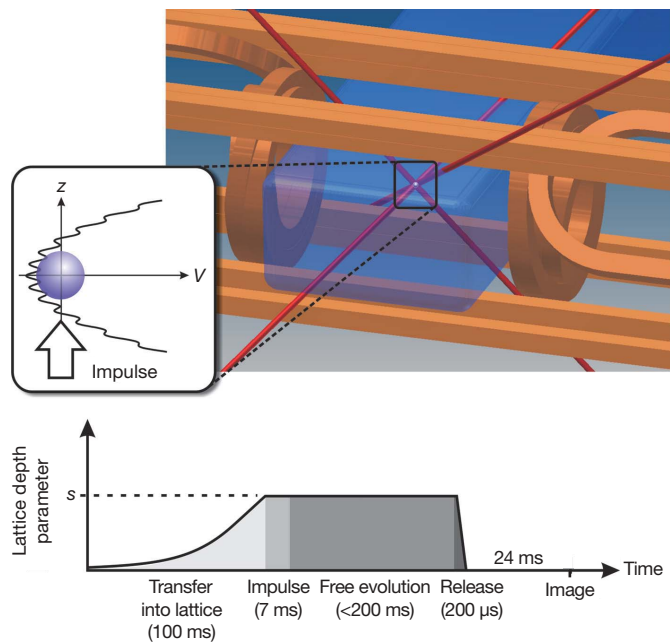
The COM motion we observe is described well by damped harmonic motion; Fig. 2 shows representative data. We fit the time evolution of the BEC COM velocity assuming the equation of motion  $m^* \ddot{z} = -2m^* \gamma \dot{z} - kz$  for the COM coordinate  $z$  in the impulse direction, where  $m^*$  is the effective mass and dots denote differentiation with respect to time. The damping rate  $\gamma$ , the oscillation frequency  $\omega = \sqrt{k/m^* - \gamma^2}$  and the initial velocity (which ranges from  $\sim 0.8$ – $1.8 \text{ mm s}^{-1}$  for the data used in this paper) are left as free parameters in the fit. Our model of the BEC COM motion assumes that the COM is acted on by the restoring force  $-kz$ , from the harmonic confinement, and a dissipative force  $-2m^* \gamma \dot{z}$ . The damping rate  $\gamma$ , which is the exponential decay rate for the amplitude of the motion, is the equivalent of electrical resistance in a solid. This can be understood by considering that ohmic resistance in a material—regardless of the source of dissipation—leads to a force on the charge carriers proportional to  $-\rho v$ , where  $\rho$  is the resistivity and  $v$  is the charge carrier velocity.

The temperature dependence of the damping rate  $\gamma$  is shown in Fig. 3. We measure the ratio of temperature  $T$  to the temperature  $T_c$  critical for condensation in the magnetic trap by determining the fraction of atoms in the BEC before transfer into the lattice. In Fig. 3 we show the values of  $\gamma$  measured for  $s = 2$  and  $s = 6$ , which correspond to lattice potential depths that sample only the superfluid region of the Bose–Hubbard phase diagram at zero temperature<sup>9</sup>. For  $s = 2$ , equation (1) requires tunnelling energy corrections beyond tight binding, which must be considered in making a detailed comparison between theory and the data shown in Fig. 3b.

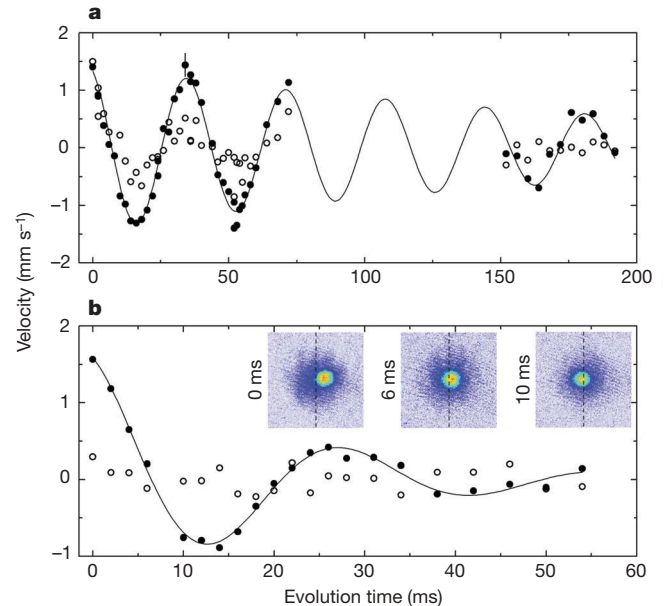
The damping rate data are plotted against inverse temperature because we fit the data to a model,  $\gamma = \gamma_0 + (\gamma_\infty - \gamma_0)e^{-\Delta E/(T/T_c)}$ , of thermally activated damping that permits a finite zero-temperature

damping rate  $\gamma_0$  ( $\gamma_\infty$  is the damping rate at infinite temperature). The data show excellent agreement with this model, excluding the highest temperature data point for each value of  $s$ . The fitted  $\gamma_0$  values, namely  $1.4 \pm 0.4 \text{ Hz}$  for  $s = 2$  and  $13 \pm 2 \text{ Hz}$  for  $s = 6$ , are inconsistent with zero at a confidence level greater than 95%. Consistency with a model that includes finite dissipation at zero temperature is, by definition, the equivalent of metallic behaviour in a solid—metals, superconductors and insulators are defined as materials that possess finite, zero and infinite resistance at zero temperature, respectively<sup>7</sup>. We have verified that the levelling off of  $\gamma$  at low temperature is not caused by saturation of  $T/T_c$  in the lattice, by measuring the fraction of atoms in the BEC after release from the lattice. This fraction does not change significantly after transfer into the lattice for the range of  $s$  and  $T/T_c$  values in Fig. 3.

The data shown in Fig. 3 are qualitatively consistent with a predicted crossover between quantum tunnelling and thermal activation of phase slips<sup>21</sup>. Phase-slip models of dissipation were first used to describe the intrinsic critical velocity of superfluid <sup>4</sup>He (ref. 1) and the appearance of resistance in a superconducting wire<sup>2</sup>. In the context of the work reported here, phase slips permit the BEC COM velocity to relax through generation of topological phase structures such as vortices and vortex rings. The COM motion of the BEC is metastable because there is an activation barrier to a phase slip event occurring and driving the system towards the zero-velocity ground state. Phase slips occur when the system tunnels through the activation barrier or when thermal fluctuations produce the required activation energy. The dissipation rate should therefore change from exhibiting thermally activated behaviour (that is, behaviour exponentially dependent on the inverse temperature) to being temperature independent at a characteristic temperature for any system in



**Figure 1 | Experimental apparatus and experimental sequence.** Over a 100-ms interval a <sup>87</sup>Rb BEC (blue sphere) is transferred from a magnetic trap (copper coloured) into an optical lattice by superimposing three pairs of counterpropagating lattice beams (red). A 7-ms impulse is applied to excite COM motion in the overall potential  $V$  experienced by the atoms, which is the combination of the lattice potential and a parabolic confining potential provided by the magnetic and optical dipole trap. After allowing up to 200 ms for the motion of the BEC to evolve, the lattice and magnetic trap are turned off and the BEC is released. The COM velocity of the BEC is measured after 24 ms of free expansion by taking an image. The inset is not to scale—the BEC diameter ranges from 30–40  $\mu\text{m}$  for the data used in this paper (corresponding to 75–100 lattice sites) and the maximum amplitude of COM motion in the harmonic potential is 10  $\mu\text{m}$ .



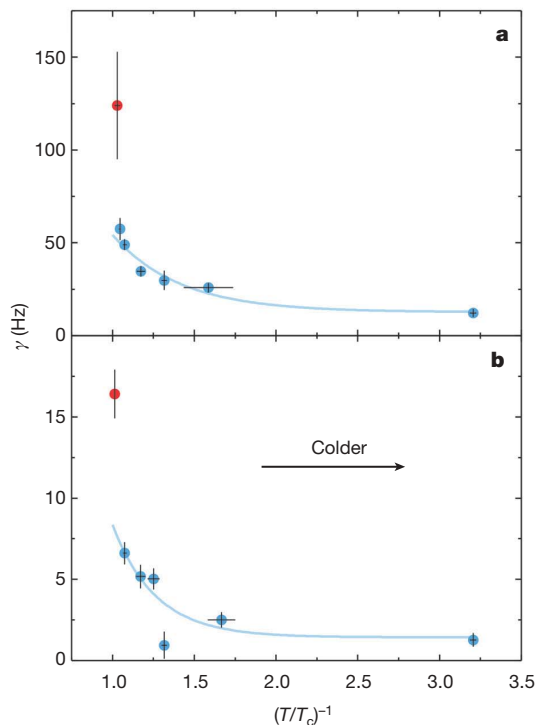
**Figure 2 | Representative data used to measure the damping rate  $\gamma$ .** Each data point represents the result of one experimental cycle, in which the COM velocities of the BEC (filled circles) and the thermal component (open circles) are measured. **a**, Data for  $s = 2$ ,  $T/T_c = 0.85$  and  $8 \times 10^5$  BEC atoms. **b**, Data for  $s = 6$ ,  $T/T_c = 0.93$  and  $2.7 \times 10^5$  BEC atoms. In **a**, the error bar is the standard deviation in the COM velocity if no impulse is applied, and the solid line is a fit to the data and is used to determine the damping rate  $\gamma$ . False-colour images that are used to measure velocity are shown as insets in **b**, and are labelled by the evolution time at which they were taken; the dashed lines bisecting the 0.47-mm field of view are there to guide the eye. Regions of high density are coloured red, and low-density regions are coloured blue. In each image, the narrow component is the BEC and the broad component is the thermal component.



which phase slips are the dominant dissipation mechanism. The data in Fig. 3 show excellent agreement with this crossover behaviour, except at the highest temperature.

Other dissipation mechanisms may contribute to the damping rate at high temperature. For example, mutual friction between the BEC and the thermal component may play a role at high temperature and for large values of  $s$ , where we observe relative motion between the BEC and the thermal component (see Fig. 2b). Mutual friction, first observed in second-sound experiments in superfluid  $^4\text{He}$  (ref. 23) and later in experiments with atomic BECs<sup>24,25</sup>, arises from interactions between atoms in the BEC and the thermal component. Because we observe the thermal component typically to be immobile at high temperature and for large values of  $s$  (similar behaviour was observed in ref. 26), mutual friction will tend to bring the BEC to rest. Mutual friction, or any dissipation mechanism induced by the thermal component, cannot explain the measured value of  $\gamma_0$ , because the thermal component vanishes at zero temperature.

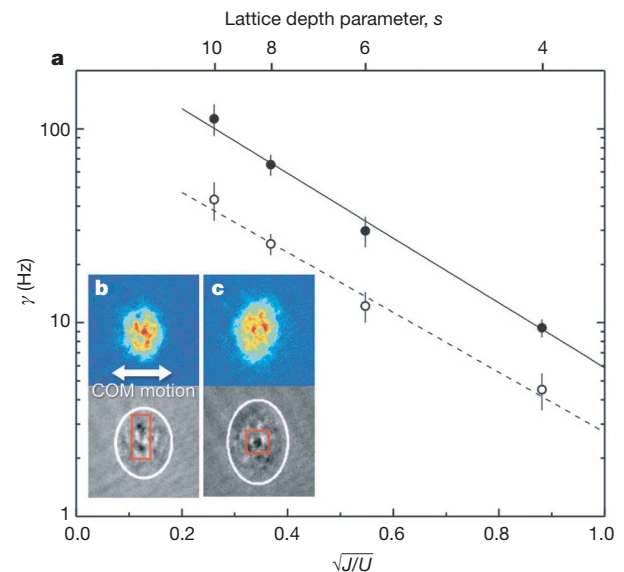
In the quantum tunnelling regime, the phase-slip rate is predicted to be proportional to  $e^{-S}$ , where  $S$  is an action characterizing the process of quantum fluctuations driving the system to lower velocity<sup>2,27</sup>. A generic scaling law for the action,  $S \propto \sqrt{J/U}$ , was derived for the Bose–Hubbard model in the appendix to ref. 21. In Fig. 4a we plot the measured damping rate for different values of  $\sqrt{J/U}$  (and thus  $S$ , up to the constant of proportionality) at two temperatures in the temperature-independent damping regime. The data fit to a line on a log–linear scale, in agreement with the relation  $\ln \gamma \propto \sqrt{J/U}$  predicted for quantum tunnelling of phase slips; the data show excellent agreement with the predicted scaling law. The systematic increase in the damping rate for the higher temperature data in Fig. 4a is consistent with a residual rate of thermally activated phase slips.



**Figure 3 | Temperature dependence of the damping rate  $\gamma$ .** **a**,  $s = 6$ ; **b**,  $s = 2$ . Each data point represents the value of  $\gamma$  determined from BEC COM motion at fixed temperature and fixed  $s$ . The red data points are excluded from the fit (solid line) to a model of thermally activated damping. The fitted energy gap  $\Delta E$  is  $k_B \times (0.5 \pm 0.1) \mu\text{K}$  for  $s = 2$  and  $k_B \times (0.33 \pm 0.08) \mu\text{K}$  for  $s = 6$  (where  $k_B$  is Boltzmann's constant). The error bars in  $\gamma$  represent the uncertainty in the fit to COM velocity data. The standard deviations in the data used to determine  $T/T_c$  for each data point are used for the error bars in  $(T/T_c)^{-1}$ .

Figure 4b, c shows features consistent with a vortex ring and a vortex directly observed in BECs undergoing COM motion. The cores of vortex rings and vortices are visible in time-of-flight images if the free expansion time is long enough for the expanded core size to be optically resolved and if the core aligns properly with the imaging axis<sup>28,29</sup>. Figure 4b, c show typical time-of-flight images in which vortex rings or vortices are observed and are recorded after 50 ms of free expansion following 2 ms of COM motion for  $s = 8$ . The features observed in these images that are consistent with a vortex ring viewed edgewise and a single vortex observed along its core are marked in red and were not detectable at the shorter expansion time used for the data in Fig. 3 and Fig. 4a. Vortex rings and single vortices nucleated by phase slips must lie in a plane perpendicular to the direction of COM motion; single vortex lines may be oriented along any direction in this plane. Vortex rings will therefore be detected edgewise by our imaging system, but single vortices will only rarely align with the imaging axis and be clearly resolved. We observe these features with approximately 20% probability for  $s = 8$ , consistent with a random generation process such as phase slips; we do not observe vortex features if COM motion is not excited. We find that vortices and vortex rings are most likely to be detected close to the edge of the BEC, where the energy gap to nucleation is smallest.

In conclusion, we observe temperature-dependent damping of COM motion in an optical lattice that is consistent with dissipation caused by phase slips. The parabolic potential used to confine the atoms gives rise to an inhomogeneous density distribution, which may enhance the effect of phase slips because the activation barrier is suppressed at the edge of the BEC. This system may therefore be comparable to thin superconducting wires and strips, in which vortices entering and leaving at the boundaries strongly influence current flow<sup>30</sup>. The effects of inhomogeneous density and finite size on phase-slip dynamics in this system remain to be conclusively theoretically addressed. The technique used in this paper can be extended to probe transport properties in models that are relevant



**Figure 4 | Scaling of  $\gamma$  with  $\sqrt{J/U}$ .** **a**, The tight-binding limit is used to calculate  $\sqrt{J/U}$  from  $s$ . The filled circles are data for  $T/T_c = 0.76$  and  $3 \times 10^5$  BEC atoms and the open circles are data for  $T/T_c = 0.312$  and  $1.1 \times 10^6$  BEC atoms. The error bars in  $\gamma$  represent the uncertainty in the fit to the COM velocity; there is an uncertainty of less than 3% in  $\sqrt{J/U}$ . The slope of  $-1.6 \pm 0.1$  determined by linear fit (solid and dashed lines) on the log–linear scale is consistent between the two temperatures. Time-of-flight images exhibiting features consistent with a vortex ring and a single vortex are shown in false colour in **b** and **c**, respectively; other excitations present in **c** are not clearly resolved. The images are fitted to a Thomas–Fermi profile, and the residual is shown in greyscale. The edge of the BEC determined from the fit is shown as a white ellipse in each inset.

to solid materials, such as two-dimensional and disordered Bose–Hubbard models. Direct imaging of vortex rings and vortices nucleated by phase slips may be used to address the microscopic dissipation dynamics in this system.

## METHODS SUMMARY

The optical lattice is created using three pairs of orthogonally polarized laser beams with wavelengths of 812 nm. These beams are weakly focused to a 120- $\mu\text{m}$  waist, and slightly frequency-offset to eliminate residual cross-dimensional interference resulting from imperfect polarization. COM motion is excited by changing the strength of a confining magnetic potential for 5 ms. The signal-to-noise ratio in measurements of the COM velocity is improved by reducing the lattice potential to zero over an interval of 200  $\mu\text{s}$ , thereby eliminating atoms effectively diffracting from the lattice potential.

The temperature is determined by measuring the fraction of atoms in the BEC. For our lowest temperature data ( $T/T_c = 0.6$  and  $T/T_c = 0.3$ ), accurately detecting the atoms in the thermal component requires images taken at high and low optical depths.

The oscillation data used to measure  $\gamma$  are checked in two ways for nonlinear response. First, no significant change in the fitted  $\gamma$  value is measured if the first period of motion is excluded from the fit. Second, fitting the data to a nonlinear damping model with  $\gamma = \Gamma|\dot{z}|^\alpha$ , where  $\alpha$  is the coefficient of nonlinear damping and  $\Gamma$  is the damping rate for  $\alpha = 0$ , typically increases the reduced  $\chi^2$  value of the fit by 10–20%. We find no clear dependence of  $\alpha$  on  $T/T_c$  or  $J/U$ , and the fitted value of  $\alpha$  averaged across all of the data is  $0.68 \pm 0.07$ . Ultimately, our sensitivity to weak velocity dependence is limited because of finite signal-to-noise ratio in measurements of the COM velocity.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 November 2007; accepted 11 March 2008.

- Langer, J. S. & Fisher, M. E. Intrinsic critical velocity of a superfluid. *Phys. Rev. Lett.* **19**, 560–563 (1967).
- Langer, J. S. & Ambegaokar, V. Intrinsic resistive transition in narrow superconducting channels. *Phys. Rev.* **164**, 498–510 (1967).
- Mooij, J. E. & Nazarov, Y. V. Superconducting nanowires as quantum phase-slip junctions. *Nature Phys.* **2**, 169–172 (2006).
- Tinkham, M. *Introduction to Superconductivity* 2nd edn, Ch. 8 (McGraw Hill, New York, 1996).
- Bezryadin, A., Lau, C. N. & Tinkham, M. Quantum suppression of superconductivity in ultrathin nanowires. *Nature* **404**, 971–974 (2000).
- Bollinger, A. T., Rogachev, A. & Bezryadin, A. Dichotomy in short superconducting nanowires: Thermal phase slippage vs. Coulomb blockade. *Europhys. Lett.* **76**, 505–511 (2006).
- Goldman, A. M. Superconductor-insulator transitions in the two-dimensional limit. *Physica E* **18**, 1–6 (2003).
- Phillips, P. & Dalidovich, D. The elusive Bose metal. *Science* **302**, 243–247 (2003).
- Fisher, M. P., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
- Greiner, M. *et al.* Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
- Jaksch, D. *et al.* Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
- Cataliotti, F. S. *et al.* Josephson junction arrays with Bose-Einstein condensates. *Science* **293**, 843–846 (2001).
- Henderson, K. *et al.* Experimental study of the role of atomic interactions on quantum transport. *Phys. Rev. Lett.* **96**, 150401 (2006).
- Fertig, C. D. *et al.* Strongly inhibited transport of a degenerate 1D Bose gas in a lattice. *Phys. Rev. Lett.* **94**, 120403 (2005).
- Cristiani, M. *et al.* Experimental properties of Bose-Einstein condensates in one-dimensional optical lattices: Bloch oscillations, Landau-Zener tunneling, and mean-field effects. *Phys. Rev. A* **65**, 063612 (2002).
- Mun, J. *et al.* Phase diagram for a Bose-Einstein condensate moving in an optical lattice. *Phys. Rev. Lett.* **99**, 150604 (2007).
- DeSarlo, L. *et al.* Unstable regimes for a Bose-Einstein condensate in an optical lattice. *Phys. Rev. A* **72**, 013603 (2005).
- Fallani, L. *et al.* Observation of dynamical instability for a Bose-Einstein condensate in a moving 1D optical lattice. *Phys. Rev. Lett.* **93**, 140406 (2004).
- Cristiani, M. *et al.* Instabilities of a Bose-Einstein condensate in a periodic potential: an experimental investigation. *Opt. Express* **12**, 4–10 (2004).
- Wu, B. & Niu, Q. Superfluidity of Bose-Einstein condensate in an optical lattice: Landau-Zener tunnelling and dynamical instability. *N. J. Phys.* **5**, 104.1–104.24 (2003).
- Polkovnikov, A. *et al.* Decay of a superfluid currents in a moving system of strongly interacting bosons. *Phys. Rev. A* **71**, 063613 (2005).
- White, M., Gao, H., Pasienski, M. & DeMarco, B. Bose-Einstein condensates in rf-dressed adiabatic potentials. *Phys. Rev. A* **74**, 023616 (2006).
- Vinen, W. F. Mutual friction in a heat current in liquid helium II. III. Theory of the mutual friction. *Proc. R. Soc. Lond. A* **242**, 493–515 (1957).
- Stamper-Kurn, D. M. *et al.* Collisionless and hydrodynamic excitations of a Bose-Einstein condensate. *Phys. Rev. Lett.* **81**, 500–503 (1998).
- Haljan, P. C., Coddington, I., Engels, P. & Cornell, E. A. Driving Bose-Einstein-condensate vorticity with a rotating normal cloud. *Phys. Rev. Lett.* **87**, 210403 (2001).
- Ferlaino, F. *et al.* Dynamics of a Bose-Einstein condensate at finite temperature in an atom-optical coherence filter. *Phys. Rev. A* **66**, 011604 (2002).
- McCumber, D. E. & Halperin, B. I. Time scale of intrinsic resistive fluctuations in thin superconducting wires. *Phys. Rev. B* **1**, 1054–1070 (1970).
- Matthews, M. R. *et al.* Vortices in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
- Anderson, B. P. *et al.* Watching dark solitons decay into vortex rings in a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2926–2929 (2001).
- Plourde, B. L. T. *et al.* Influence of edge barriers on vortex dynamics in thin weak-pinning superconducting strips. *Phys. Rev. B* **64**, 014503 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Barankov, E. Demler, P. Goldbart, N. Goldenfeld, D. Pekker and P. Phillips for discussions, and D. S. Jin, N. Mason and J. V. Porto for critically reading this manuscript. This work was supported by the National Science Foundation, the Office of Naval Research and the UIUC Research Board. D.M. acknowledges support from the Carver Foundation and NSERC.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.D. ([bdemarco@uiuc.edu](mailto:bdemarco@uiuc.edu)).



## METHODS

**Optical lattice apparatus.**  $^{87}\text{Rb}$  BECs are created in the  $|F=1, m_F=-1\rangle$  state in a magnetic trap with a 21.7-Hz radial frequency and 14.5-Hz axial frequency. The optical lattice potential

$$-s(\cos^2(\pi\hat{\mathbf{n}}_1 \cdot \mathbf{r}/d) + \cos^2(\pi\hat{\mathbf{n}}_2 \cdot \mathbf{r}/d) + \cos^2(\pi\hat{\mathbf{n}}_3 \cdot \mathbf{r}/d))$$

is formed from three retro-reflected laser beams with orthogonal wavevectors and polarizations along the directions  $\hat{\mathbf{n}}_1$ ,  $\hat{\mathbf{n}}_2$  and  $\hat{\mathbf{n}}_3$ . The beam directions make  $60^\circ$  and  $45^\circ$  angles with the vertical axis  $z$ . Two of the optical lattice beams are respectively frequency-offset by +19.66 and -176.94 MHz from the third to prevent residual cross-dimensional interference. The magnetic trap is left on during and after transfer into the lattice, to support the atoms against gravity. The impulse used to excite COM motion is created by briefly changing the equilibrium position of the BEC in the magnetic trap. The harmonic potential is not spherically symmetric as was specified for simplicity in equation (1). The percentage difference between the weakest and the geometric mean of the oscillator frequencies along the lattice directions is 8% for  $s=2$ , decreasing to 2% for  $s=6$ . For  $s=2$ , equation (1) also requires corrections for next-nearest-neighbour tunnelling.

**COM velocity measurement.** Standard absorption imaging is used to measure the COM position of the atoms, which is translated into COM velocity, after release from the optical lattice and the harmonic potential. Before turning off the magnetic trap, the lattice is linearly ramped down over a 200- $\mu\text{s}$  interval to improve the image signal-to-noise ratio. We checked that this 'band-mapping'<sup>31</sup> step does not affect the measured COM velocity, by comparing our data with some for which the lattice potential was turned off over 350 ns. Images are fitted to a Thomas–Fermi profile if only a BEC is visible, or the combination of a Thomas–Fermi and a gaussian profile if atoms in a thermal component are present. Although the COM of each component is left as a free parameter in the fit, we do not analyse motion of the thermal component in this paper.

**Temperature measurement.** We determine temperature for  $T/T_c > 0.6$  by measuring the fraction of atoms in the BEC using a single absorption image taken after the BEC's release from the magnetic trap. The atoms are optically pumped (or 'repumped') to the  $|F=2, m_F=-2\rangle$  state before imaging. The thermal fraction at our lowest temperatures is not visible for images in which the maximum optical depth is small enough to accurately measure the number of atoms in the BEC. Therefore, the condensate fractions for  $T/T_c = 0.312$  and  $T/T_c = 0.6$  are determined by taking two images. One image is of partially repumped atoms, and is used to determine the total optical depth of the BEC. The second image is of fully repumped atoms, and is used to determine the total optical depth of the thermal component by fitting the low-optical-depth region of the image. We compare the optical depths determined from the two images by calibrating the

fractional change in optical depth for partial repumping. The calibration is performed using a thermal gas and taking images at two free-expansion times.

We measure  $T/T_c$  using the fraction of atoms in the BEC because there are no proven techniques for determining  $T$  in an optical lattice. Interactions may change  $T/T_c$  for BECs loaded into the lattice for the highest value of  $s$  in Fig. 4a (refs 32–34); the extent to which this effect plays a role in experiments has not been resolved. The temperature of the gas was controlled by altering the evaporative cooling procedure, resulting, for fixed  $s$ , in a varying number of atoms in the BEC and a varying  $T_c$  value. The average  $T_c$  value for the data used in this paper is 0.13  $\mu\text{K}$ , determined from the total number of atoms and the magnetic trap oscillator frequency;  $T_c$  spans a 0.07- $\mu\text{K}$  range for the data in Fig. 3. The heating caused by the dissipation observed in our experiment cannot be detected within our experimental uncertainty in condensate fraction or absolute temperature.

**Technical noise.** We rule out several technical noise sources as dissipation mechanisms that could explain the damping of COM motion. Anharmonicity in the dipole potential may effectively damp COM motion for large values of  $s$ . To check for anharmonic behaviour, we measure COM motion when the retro-reflected lattice laser beams are removed, which eliminates the lattice potential and reduces the depth of the dipole potential by a factor of  $\sim 2$ . Using this technique we measure a value of  $\gamma$  consistent with zero for lattice laser intensities corresponding to  $s=9$  (or  $s=18$  if the retro-reflected beams are present), eliminating trap anharmonicity as an effective dissipation source for the data in Fig. 3. Relative motion between the lattice and the harmonic potential or fluctuations in  $s$  (caused by retro-reflecting mirror motion and lattice laser intensity fluctuations, respectively) can lead to the dephasing of dipole mode motion by transferring atoms into states with different values of  $m^*$  in excited bands<sup>35</sup>; we do not, however, observe population outside of the lowest energy band. The total spontaneous emission rate per atom is less than 0.3 Hz for  $s=6$ , so momentum diffusion caused by scattering light from the optical lattice laser beams is insignificant. The lattice depth varies by less than 3% across the BEC, so spatial variation in the effective mass can play no role in the dissipation timescales measured in our data.

31. Greiner, M. *et al.* Exploring phase coherence in a 2D lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160405 (2001).
32. Blakie, P. B. & Porto, J. V. Adiabatic loading of bosons into optical lattices. *Phys. Rev. A* **69**, 013603 (2004).
33. Ho, T.-L. & Zhou, Q. Intrinsic heating and cooling in adiabatic processes for bosons in optical lattices. *Phys. Rev. Lett.* **99**, 120404 (2007).
34. Rey, A., Pupillo, G. & Porto, J. V. The role of interactions, tunneling, and harmonic confinement on the adiabatic loading of bosons in an optical lattice. *Phys. Rev. A* **73**, 023608 (2006).
35. Denschlag, J. H. *et al.* A Bose-Einstein condensate in an optical lattice. *J. Phys. B* **35**, 3095–3110 (2002).

## LETTERS

## The missing memristor found

Dmitri B. Strukov<sup>1</sup>, Gregory S. Snider<sup>1</sup>, Duncan R. Stewart<sup>1</sup> & R. Stanley Williams<sup>1</sup>

Anyone who ever took an electronics laboratory class will be familiar with the fundamental passive circuit elements: the resistor, the capacitor and the inductor. However, in 1971 Leon Chua reasoned from symmetry arguments that there should be a fourth fundamental element, which he called a memristor (short for memory resistor)<sup>1</sup>. Although he showed that such an element has many interesting and valuable circuit properties, until now no one has presented either a useful physical model or an example of a memristor. Here we show, using a simple analytical example, that memristance arises naturally in nanoscale systems in which solid-state electronic and ionic transport are coupled under an external bias voltage. These results serve as the foundation for understanding a wide range of hysteretic current–voltage behaviour observed in many nanoscale electronic devices<sup>2–19</sup> that involve the motion of charged atomic or molecular species, in particular certain titanium dioxide cross-point switches<sup>20–22</sup>.

More specifically, Chua noted that there are six different mathematical relations connecting pairs of the four fundamental circuit variables: electric current  $i$ , voltage  $v$ , charge  $q$  and magnetic flux  $\phi$ . One of these relations (the charge is the time integral of the current) is determined from the definitions of two of the variables, and another (the flux is the time integral of the electromotive force, or voltage) is determined from Faraday's law of induction. Thus, there should be four basic circuit elements described by the remaining relations between the variables (Fig. 1). The 'missing' element—the memristor, with memristance  $M$ —provides a functional relation between charge and flux,  $d\phi = M dq$ .

In the case of linear elements, in which  $M$  is a constant, memristance is identical to resistance and, thus, is of no special interest. However, if  $M$  is itself a function of  $q$ , yielding a nonlinear circuit element, then the situation is more interesting. The  $i$ – $v$  characteristic of such a nonlinear relation between  $q$  and  $\phi$  for a sinusoidal input is generally a frequency-dependent Lissajous figure<sup>1</sup>, and no combination of nonlinear resistive, capacitive and inductive components can duplicate the circuit properties of a nonlinear memristor (although including active circuit elements such as amplifiers can do so)<sup>1</sup>. Because most valuable circuit functions are attributable to nonlinear device characteristics, memristors compatible with integrated circuits could provide new circuit functions such as electronic resistance switching at extremely high two-terminal device densities. However, until now there has not been a material realization of a memristor.

The most basic mathematical definition of a current-controlled memristor for circuit analysis is the differential form

$$v = \mathcal{R}(w)i \quad (1)$$

$$\frac{dw}{dt} = i \quad (2)$$

where  $w$  is the state variable of the device and  $\mathcal{R}$  is a generalized resistance that depends upon the internal state of the device. In this case the state variable is just the charge, but no one has been able to

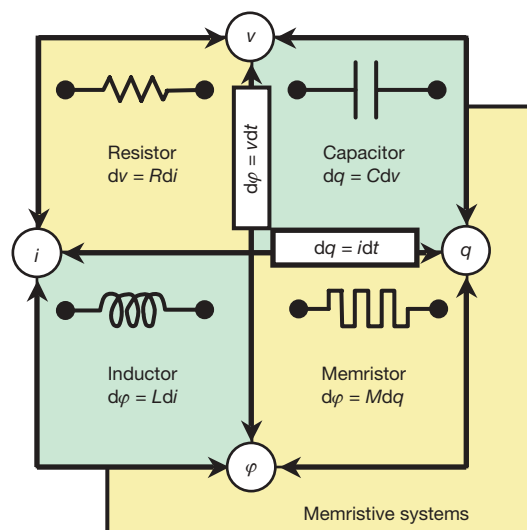
propose a physical model that satisfies these simple equations. In 1976 Chua and Kang generalized the memristor concept to a much broader class of nonlinear dynamical systems they called memristive systems<sup>23</sup>, described by the equations

$$v = \mathcal{R}(w,i)i \quad (3)$$

$$\frac{dw}{dt} = f(w,i) \quad (4)$$

where  $w$  can be a set of state variables and  $\mathcal{R}$  and  $f$  can in general be explicit functions of time. Here, for simplicity, we restrict the discussion to current-controlled, time-invariant, one-port devices. Note that, unlike in a memristor, the flux in memristive systems is no longer uniquely defined by the charge. However, equation (3) does serve to distinguish a memristive system from an arbitrary dynamical device; no current flows through the memristive system when the voltage drop across it is zero. Chua and Kang showed that the  $i$ – $v$  characteristics of some devices and systems, notably thermistors, Josephson junctions, neon bulbs and even the Hodgkin–Huxley model of the neuron, can be modelled using memristive equations<sup>23</sup>. Nevertheless, there was no direct connection between the mathematics and the physical properties of any practical system, and hence, almost forty years later, the concepts have not been widely adopted.

Here we present a physical model of a two-terminal electrical device that behaves like a perfect memristor for a certain restricted



**Figure 1 | The four fundamental two-terminal circuit elements: resistor, capacitor, inductor and memristor.** Resistors and memristors are subsets of a more general class of dynamical devices, memristive systems. Note that  $R$ ,  $C$ ,  $L$  and  $M$  can be functions of the independent variable in their defining equations, yielding nonlinear elements. For example, a charge-controlled memristor is defined by a single-valued function  $M(q)$ .

<sup>1</sup>HP Labs, 1501 Page Mill Road, Palo Alto, California 94304, USA.



range of the state variable  $w$  and as a memristive system for another, wider (but still bounded), range of  $w$ . This intuitive model produces rich hysteretic behaviour controlled by the intrinsic nonlinearity of  $M$  and the boundary conditions on the state variable  $w$ . The results provide a simplified explanation for reports of current–voltage anomalies, including switching and hysteretic conductance, multiple conductance states and apparent negative differential resistance, especially in thin-film, two-terminal nanoscale devices, that have been appearing in the literature for nearly 50 years<sup>2–4</sup>.

Electrical switching in thin-film devices has recently attracted renewed attention, because such a technology may enable functional scaling of logic and memory circuits well beyond the limits of complementary metal-oxide-semiconductors<sup>24,25</sup>. The microscopic nature of resistance switching and charge transport in such devices is still under debate, but one proposal is that the hysteresis requires some sort of atomic rearrangement that modulates the electronic current. On the basis of this proposition, we consider a thin semiconductor film of thickness  $D$  sandwiched between two metal contacts, as shown in Fig. 2a. The total resistance of the device is determined by two variable resistors connected in series (Fig. 2a), where the resistances are given for the full length  $D$  of the device. Specifically, the semiconductor film has a region with a high concentration of dopants (in this example assumed to be positive ions) having low resistance  $\mathcal{R}_{\text{ON}}$ , and the remainder has a low (essentially zero) dopant concentration and much higher resistance  $\mathcal{R}_{\text{OFF}}$ .

The application of an external bias  $v(t)$  across the device will move the boundary between the two regions by causing the charged dopants to drift<sup>26</sup>. For the simplest case of ohmic electronic conduction and linear ionic drift in a uniform field with average ion mobility

$\mu_V$ , we obtain

$$v(t) = \left( \mathcal{R}_{\text{ON}} \frac{w(t)}{D} + \mathcal{R}_{\text{OFF}} \left( 1 - \frac{w(t)}{D} \right) \right) i(t) \quad (5)$$

$$\frac{dw(t)}{dt} = \mu_V \frac{\mathcal{R}_{\text{ON}}}{D} i(t) \quad (6)$$

which yields the following formula for  $w(t)$ :

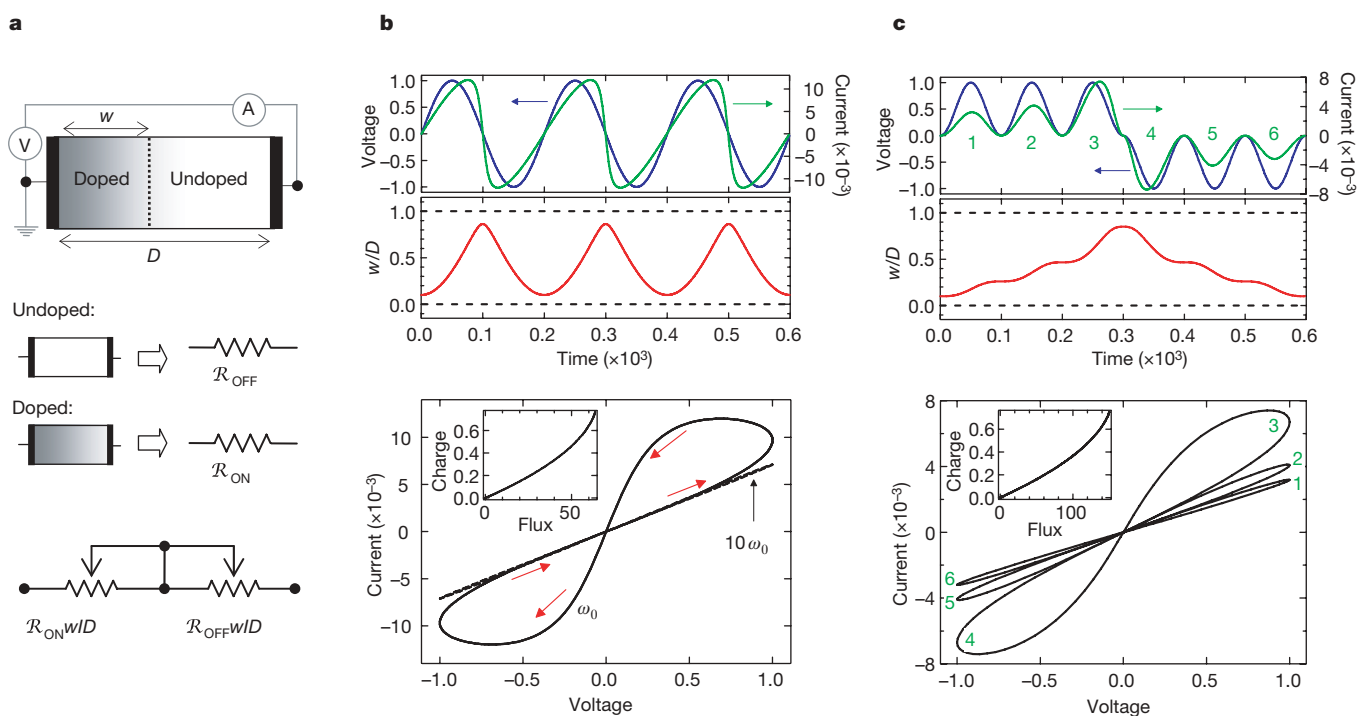
$$w(t) = \mu_v \frac{\mathcal{R}_{\text{ON}}}{D} q(t) \quad (7)$$

By inserting equation (7) into equation (5) we obtain the memristance of this system, which for  $\mathcal{R}_{\text{ON}} \ll \mathcal{R}_{\text{OFF}}$  simplifies to:

$$M(q) = \mathcal{R}_{\text{OFF}} \left( 1 - \frac{\mu_V \mathcal{R}_{\text{ON}}}{D^2} q(t) \right)$$

The  $q$ -dependent term in parentheses on the right-hand side of this equation is the crucial contribution to the memristance, and it becomes larger in absolute value for higher dopant mobilities  $\mu_V$  and smaller semiconductor film thicknesses  $D$ . For any material, this term is 1,000,000 times larger in absolute value at the nanometre scale than it is at the micrometre scale, because of the factor of  $1/D^2$ , and the memristance is correspondingly more significant. Thus, memristance becomes more important for understanding the electronic characteristics of any device as the critical dimensions shrink to the nanometre scale.

The coupled equations of motion for the charged dopants and the electrons in this system take the normal form for a current-controlled (or charge-controlled) memristor (equations (1) and (2)). The fact that the magnetic field does not play an explicit role in the



**Figure 2 | The coupled variable-resistor model for a memristor.** **a**, Diagram with a simplified equivalent circuit. V, voltmeter; A, ammeter. **b**, **c**, The applied voltage (blue) and resulting current (green) as a function of time  $t$  for a typical memristor. In **b** the applied voltage is  $v_0 \sin(\omega_0 t)$  and the resistance ratio is  $R_{\text{OFF}}/R_{\text{ON}} = 160$ , and in **c** the applied voltage is  $\pm v_0 \sin^2(\omega_0 t)$  and  $R_{\text{OFF}}/R_{\text{ON}} = 380$ , where  $v_0$  is the magnitude of the applied voltage and  $\omega_0$  is the frequency. The numbers 1–6 label successive waves in the applied voltage and the corresponding loops in the  $i$ - $v$  curves. In each plot the axes are dimensionless, with voltage, current, time, flux and charge expressed in units of  $v_0 = 1$  V,  $i_0 \equiv v_0/R_{\text{ON}} = 10$  mA,  $t_0 \equiv 2\pi/\omega_0 \equiv D^2/\mu_{\text{V}}v_0 = 10$  ms,  $v_0 t_0$  and

$i_0 t_0$ , respectively. Here  $i_0$  denotes the maximum possible current through the device, and  $t_0$  is the shortest time required for linear drift of dopants across the full device length in a uniform field  $v_0/D$ , for example with  $D = 10$  nm and  $\mu_V = 10^{-10}$  cm<sup>2</sup> s<sup>-1</sup> V<sup>-1</sup>. We note that, for the parameters chosen, the applied bias never forces either of the two resistive regions to collapse; for example,  $w/D$  does not approach zero or one (shown with dashed lines in the middle plots in **b** and **c**). Also, the dashed  $i$ - $v$  plot in **b** demonstrates the hysteresis collapse observed with a tenfold increase in sweep frequency. The insets in the  $i$ - $v$  plots in **b** and **c** show that for these examples the charge is a single-valued function of the flux, as it must be in a memristor.

mechanism of memristance is one possible reason why the phenomenon has been hidden for so long; those interested in memristive devices were searching in the wrong places. The mathematics simply require there to be a nonlinear relationship between the integrals of the current and voltage, which is realized in equations (5) and (6). Another significant issue that was not anticipated by Chua is that the state variable  $w$ , which in this case specifies the distribution of dopants in the device, is bounded between zero and  $D$ . The state variable is proportional to the charge  $q$  that passes through the device until its value approaches  $D$ ; this is the condition of 'hard' switching (large voltage excursions or long times under bias). As long as the system remains in the memristor regime, any symmetrical alternating-current voltage bias results in double-loop  $i$ - $v$  hysteresis that collapses to a straight line for high frequencies (Fig. 2b). Multiple continuous states will also be obtained if there is any sort of asymmetry in the applied bias (Fig. 2c).

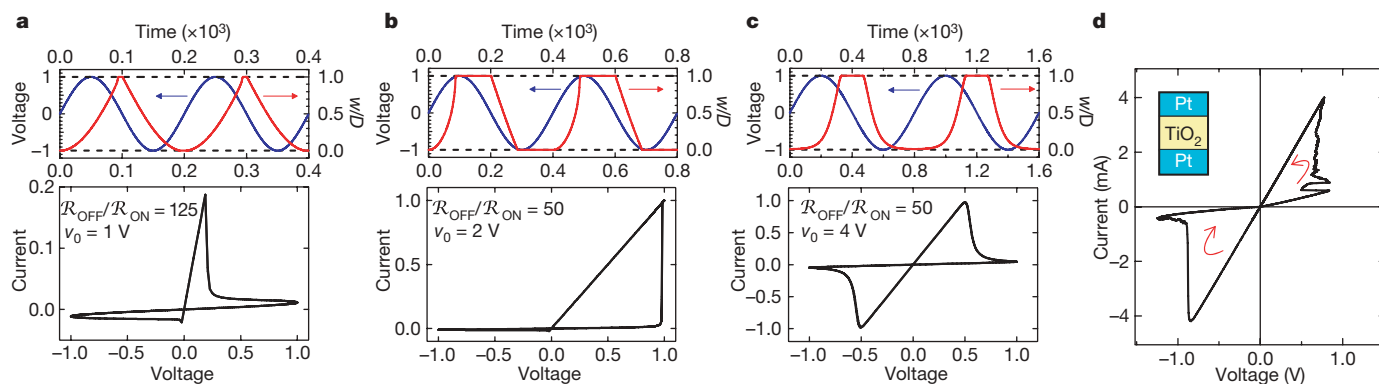
Obviously, equation (7) is only valid for values of  $w$  in the interval  $[0, D]$ . Different hard-switching cases are defined by imposing a variety of boundary conditions, such as assuming that once the value of  $w$  reaches either of the boundaries, it remains constant until the voltage reverses polarity. In such a case, the device satisfies the normal equations for a current-controlled memristive system (equations (3) and (4)). Figure 3a, b shows two qualitatively different  $i$ - $v$  curves that are possible for such a memristive device. In Fig. 3a, the upper boundary is reached while the derivative of the voltage is negative, producing an apparent or 'dynamical' negative differential resistance. Unlike a true 'static' negative differential resistance, which would be insensitive to time and device history, such a dynamical effect is simply a result of the charge-dependent change in the device resistance, and can be identified by a strong dependence on the frequency of a sinusoidal driving voltage. In another case, for example when the boundary is reached much faster by doubling the magnitude of the applied voltage (Fig. 3b), the switching event is a monotonic function of current. Even though in the hard-switching case there appears to be a clearly defined threshold voltage for switching from the 'off' (high resistance) state to the 'on' (low resistance) state, the effect is actually dynamical. This means that any positive voltage  $v_+$  applied to the device in the off state will eventually switch it to the on state after time  $\sim D^2 R_{\text{OFF}} / (2\mu_V v_+ R_{\text{ON}})$ . The device will remain in the on state as long as a positive voltage is applied, but even a small negative bias will switch it back to the off state; this is why a current-hysteresis loop is only observed for the positive voltage sweep in Fig. 3a, b.

In nanoscale devices, small voltages can yield enormous electric fields, which in turn can produce significant nonlinearities in ionic transport. Figure 3c illustrates such a case in which the right-hand side of equation (6) is multiplied by a window function  $w(1-w)/D^2$ , which corresponds to nonlinear drift when  $w$  is close to zero or  $D$ . In

this case, the switching event requires a significantly larger amount of charge (or even a threshold voltage) in order for  $w$  to approach either boundary. Therefore, the switching is essentially binary because the on and off states can be held much longer if the voltage does not exceed a specific threshold. Nonlinearity can also be expected in the electronic transport, which can be due to, for example, tunnelling at the interfaces or high-field electron hopping. In this case, the hysteresis behaviour discussed above remains essentially the same but the  $i$ - $v$  characteristic becomes nonlinear.

The model of equations (5) and (6) exhibits many features that have been described as bipolar switching, that is, when voltages of opposite polarity are required for switching a device to the on state and the off state. This type of behaviour has been experimentally observed in various material systems: organic films<sup>5-9</sup> that contain charged dopants or molecules with mobile charged components; chalcogenides<sup>4,10-12</sup>, where switching is attributed to ion migration rather than a phase transition; and metal oxides<sup>2-4,20</sup>, notably  $\text{TiO}_2$  (refs 4, 13, 14, 21) and various perovskites<sup>4,15-19</sup>. For example, multi-state<sup>8-14,16-18,20,21</sup> and binary<sup>3,4,7,15,16</sup> switching that are similar to those modelled in Figs 2c and 3c, respectively, have been observed, with some showing dynamical negative differential resistance. Typically, hysteresis such as in Fig. 3c is observed for both voltage polarities<sup>7,9-12,14-17,21</sup>, but observations of  $i$ - $v$  characteristics resembling Fig. 3a, b have also been reported<sup>8,17-20</sup>. In our own studies of  $\text{TiO}_x$  devices,  $i$ - $v$  behaviours very similar to those in Figs 2b, 2c and 3c are regularly observed. Figure 3d illustrates an experimental  $i$ - $v$  characteristic from a metal/oxide/metal cross-point device within which the critical 5-nm-thick oxide film initially contained one layer of insulating  $\text{TiO}_2$  and one layer of oxygen-poor  $\text{TiO}_{2-x}$  (refs 21, 22). In this system, oxygen vacancies act as mobile +2-charged dopants, which drift in the applied electric field, shifting the dividing line between the  $\text{TiO}_2$  and  $\text{TiO}_{2-x}$  layers. The switching characteristic observed for a particular memristive system helps classify the nature of the boundary conditions on the state variable of the device.

The rich hysteretic  $i$ - $v$  characteristics detected in many thin-film, two-terminal devices can now be understood as memristive behaviour defined by coupled equations of motion: some for (ionized) atomic degrees of freedom that define the internal state of the device, and others for the electronic transport. This behaviour is increasingly relevant as the active region in many electronic devices continues to shrink to a width of only a few nanometres, so even a low applied voltage corresponds to a large electric field that can cause charged species to move. Such dopant or impurity motion through the active region can produce dramatic changes in the device resistance. Including memristors and memristive systems in integrated circuits has the potential to significantly extend circuit functionality as long as the dynamical nature of such devices is understood and properly



**Figure 3 | Simulations of a voltage-driven memristive device.** **a**, Simulation with dynamic negative differential resistance; **b**, simulation with no dynamic negative differential resistance; **c**, simulation governed by nonlinear ionic drift. In the upper plots of **a**, **b** and **c** we plot the voltage stimulus (blue) and the corresponding change in the normalized state variable  $w/D$  (red), versus

time. In all cases, hard switching occurs when  $w/D$  closely approaches the boundaries at zero and one (dashed), and the qualitatively different  $i$ - $v$  hysteresis shapes are due to the specific dependence of  $w/D$  on the electric field near the boundaries. **d**, For comparison, we present an experimental  $i$ - $v$  plot of a Pt- $\text{TiO}_{2-x}$ -Pt device<sup>21</sup>.



used. Important applications include ultradense, semi-non-volatile memories and learning networks that require a synapse-like function.

Received 6 December 2007; accepted 17 March 2008.

- Chua, L. O. Memristor - the missing circuit element. *IEEE Trans. Circuit Theory* **18**, 507–519 (1971).
- Hickmott, M. T. Low-frequency negative resistance in thin anodic oxide films. *J. Appl. Phys.* **33**, 2669–2682 (1962).
- Dearnaley, G., Stoneham, A. M. & Morgan, D. V. Electrical phenomena in amorphous oxide films. *Rep. Prog. Phys.* **33**, 1129–1192 (1970).
- Waser, R. & Aono, M. Nanoionics-based resistive switching memories. *Nature Mater.* **6**, 833–840 (2007).
- Scott, J. C. & Bozano, L. D. Nonvolatile memory elements based on organic materials. *Adv. Mater.* **19**, 1452–1463 (2007).
- Collier, C. P. *et al.* A [2]catenane-based solid state electronically reconfigurable switch. *Science* **289**, 1172–1175 (2000).
- Zhitenev, N. B., Sidorenko, A., Tennant, D. M. & Cirelli, R. A. Chemical modification of the electronic conducting states in polymer nanodevices. *Nature Nanotechnol.* **2**, 237–242 (2007).
- Smits, J. H. A., Meskers, S. C. J., Janssen, R. A. J., Marsman, A. W. & de Leeuw, D. M. Electrically rewritable memory cells from poly(3-hexylthiophene) Schottky diodes. *Adv. Mater.* **17**, 1169–1173 (2005).
- Lai, Q. X., Zhu, Z. H., Chen, Y., Patil, S. & Wudl, F. Organic nonvolatile memory by dopant-configurable polymer. *Appl. Phys. Lett.* **88**, 133515 (2006).
- Terabe, K., Hasegawa, T., Nakayama, T. & Aono, M. Quantized conductance atomic switch. *Nature* **433**, 47–50 (2005).
- Kozicki, M. N., Park, M. & Mitkova, M. Nanoscale memory elements based on solid-state electrolytes. *IEEE Trans. Nanotechnol.* **4**, 331–338 (2005).
- Dietrich, S. *et al.* A nonvolatile 2-Mbit CBRAM memory core featuring advanced read and program control. *IEEE J. Solid State Circuits* **42**, 839–845 (2007).
- Jameson, J. R. *et al.* Field-programmable rectification in rutile TiO<sub>2</sub> crystals. *Appl. Phys. Lett.* **91**, 112101 (2007).
- Jeong, D. S., Schroeder, H. & Waser, R. Coexistence of bipolar and unipolar resistive switching behaviors in a Pt/TiO<sub>2</sub>/Pt stack. *Electrochem. Solid State Lett.* **10**, G51–G53 (2007).
- Beck, A., Bednorz, J. G., Gerber, C., Rossel, C. & Widmer, D. Reproducible switching effect in thin oxide films for memory applications. *Appl. Phys. Lett.* **77**, 139–141 (2000).
- Szot, K., Speier, W., Bihlmayer, G. & Waser, R. Switching the electrical resistance of individual dislocations in single-crystalline SrTiO<sub>3</sub>. *Nature Mater.* **5**, 312–320 (2006).
- Sawa, A., Fujii, T., Kawasaki, M. & Tokura, Y. Interface resistance switching at a few nanometer thick perovskite manganite active layers. *Appl. Phys. Lett.* **88**, 232112 (2006).
- Hamaguchi, M., Aoyama, K., Asanuma, S., Uesu, Y. & Katsufuji, T. Electric-field-induced resistance switching universally observed in transition-metal-oxide thin films. *Appl. Phys. Lett.* **88**, 142508 (2006).
- Oligschlaeger, R., Waser, R., Meyer, R., Karthäuser, S. & Dittmann, R. Resistive switching and data reliability of epitaxial (Ba,Sr)TiO<sub>3</sub> thin films. *Appl. Phys. Lett.* **88**, 042901 (2006).
- Richter, C. A., Stewart, D. R., Ohlberg, D. A. A. & Williams, R. S. Electrical characterization of Al/AIO<sub>x</sub>/molecule/Ti/Al devices. *Appl. Phys. Mater. Sci. Process.* **80**, 1355–1362 (2005).
- Stewart, D. R. *et al.* Molecule-independent electrical switching in Pt/organic monolayer/Ti devices. *Nano Lett.* **4**, 133–136 (2004).
- Blackstock, J. J., Stickle, W. F., Donley, C. L., Stewart, D. R. & Williams, R. S. Internal structure of a molecular junction device: chemical reduction of PtO<sub>2</sub> by Ti evaporation onto an interceding organic monolayer. *J. Phys. Chem. C* **111**, 16–20 (2007).
- Chua, L. O. & Kang, S. M. Memristive devices and systems. *Proc. IEEE* **64**, 209–223 (1976).
- Kuekes, P. J., Snider, G. S. & Williams, R. S. Crossbar nanocomputers. *Sci. Am.* **293**, 72–78 (2005).
- Strukov, D. B. & Likharev, K. K. Defect-tolerant architectures for nanoelectronic crossbar memories. *J. Nanosci. Nanotechnol.* **7**, 151–167 (2007).
- Blanc, J. & Staebler, D. L. Electrocoloration in SrTiO - vacancy drift and oxidation-reduction of transition metals. *Phys. Rev. B* **4**, 3548–3557 (1971).

**Acknowledgements** This research was conducted with partial support from DARPA and DTO.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.S.W. (stan.williams@hp.com).

## CORRIGENDUM

doi:10.1038/nature08166

**The missing memristor found**

Dmitri B. Strukov, Gregory S. Snider, Duncan R. Stewart  
& R. Stanley Williams

*Nature* 453, 80–83 (2008)

---

In Fig. 2a of this Letter, the resistance for the right-hand element of the bottom circuit should be  $R_{\text{OFF}}(1 - w/D)$ , instead of the shown value  $R_{\text{OFF}}w/D$ . Also, the correct value of window function should be  $w(D - W)/D^2$ , rather than  $w(1 - W)/D^2$ . All simulation results presented in the paper use the correct formulae.

## LETTERS

# Advancing decadal-scale climate prediction in the North Atlantic sector

N. S. Keenlyside<sup>1</sup>, M. Latif<sup>1</sup>, J. Jungclauss<sup>2</sup>, L. Kornblueh<sup>2</sup> & E. Roeckner<sup>2</sup>

The climate of the North Atlantic region exhibits fluctuations on decadal timescales that have large societal consequences. Prominent examples include hurricane activity in the Atlantic<sup>1</sup>, and surface-temperature and rainfall variations over North America<sup>2</sup>, Europe<sup>3</sup> and northern Africa<sup>4</sup>. Although these multidecadal variations are potentially predictable if the current state of the ocean is known<sup>5–7</sup>, the lack of subsurface ocean observations<sup>8</sup> that constrain this state has been a limiting factor for realizing the full skill potential of such predictions<sup>9</sup>. Here we apply a simple approach—that uses only sea surface temperature (SST) observations—to partly overcome this difficulty and perform retrospective decadal predictions with a climate model. Skill is improved significantly relative to predictions made with incomplete knowledge of the ocean state<sup>10</sup>, particularly in the North Atlantic and tropical Pacific oceans. Thus these results point towards the possibility of routine decadal climate predictions. Using this method, and by considering both internal natural climate variations and projected future anthropogenic forcing, we make the following forecast: over the next decade, the current Atlantic meridional overturning circulation will weaken to its long-term mean; moreover, North Atlantic SST and European and North American surface temperatures will cool slightly, whereas tropical Pacific SST will remain almost unchanged. Our results suggest that global surface temperature may not increase over the next decade, as natural climate variations in the North Atlantic and tropical Pacific temporarily offset the projected anthropogenic warming.

The North Atlantic is a region with large natural multidecadal variability<sup>2,4,11–14</sup>. Observations suggest that Atlantic multidecadal variability may be oscillatory, with a period of 70–80 years. Palaeo-evidence<sup>15</sup> and coupled general circulation models<sup>16–18</sup> support the oscillatory nature of Atlantic multidecadal variability, and hence the existence of a low order mode of natural variability. Although major uncertainties exist in the mechanisms of Atlantic multidecadal variability<sup>19</sup>, it is widely accepted that the meridional overturning circulation (MOC) plays an important role in driving multidecadal SST variations, as shown by ocean<sup>20</sup> and coupled<sup>5,7,14,16,19</sup> model simulations.

Studies with coupled general circulation models assuming perfect ocean initial conditions indicate that accurate initialization of MOC may allow Atlantic multidecadal variability to be predicted a decade or more in advance<sup>5–7</sup>. However, past MOC fluctuations have been poorly observed<sup>8</sup> and large uncertainties exist among ocean model simulations<sup>21</sup> and ocean analyses. Here, this difficulty is partly overcome with a simple initialization scheme, previously applied in seasonal forecasting<sup>22</sup>, that consists of relaxing SST anomalies of the coupled general circulation models to observations. (These simulations, called SST-restored, are described in Methods.) Additionally, decadal predictions require accurate projections of external radiative

forcing (that is, computed from concentrations of greenhouse gases and sulphate aerosols, solar cycle variations, and volcanic activity)<sup>10,23</sup>.

Using this simple initialization technique and a state-of-the-art coupled general circulation model<sup>24</sup>, a set of ten-year-long, three-member ensemble hindcasts/forecasts were performed every five years from 1955 till 2005 (except for 1995, which instead is started in 1994). External radiative forcing in the hindcasts is treated following ref. 9. Three additional integrations with radiative forcing computed following observations (called twentieth century-RF) are used to estimate predictability due to external radiative forcing. (See Methods for experimental details.)

Skilful predictions of surface temperature averaged over years one to ten of the hindcasts are obtained over large parts of the North Atlantic, Europe, North America and northern Africa (Fig. 1a). In these regions, skill is well above that of persistence (Fig. 1b). Persistence, the prediction that assumes no change from the initial conditions, is a commonly used benchmark for skill. The twentieth century-RF simulations show skill over parts of northern Africa, North America and Europe, but no skill over the North Atlantic Ocean, except in the vicinity of the Equator (Fig. 1c). Over the northern North Atlantic, the climate model hindcast is significantly more skilful than the twentieth century-RF simulations (Fig. 1a). It follows that in this region a significant fraction of the skill arises from initialization of the North Atlantic Ocean, as opposed to external radiative forcing. Over western Europe and large parts of North America, initialization also leads to a significant enhancement in skill. Skill enhancements outside the North Atlantic sector are found in the central eastern tropical Pacific (Fig. 1a and below) and the equatorial Indian Ocean (Supplementary Information).

In several regions, initialization causes a significant degradation in skill compared to the twentieth century-RF simulations (Fig. 1c). Over the tropical North Atlantic and central Africa this is probably related to deficiencies, common to many coupled models, in simulating tropical Atlantic climate. Thus, large skill improvements in this region may be expected, as models improve. The SST-restored simulations (Fig. 1d) provide an estimate for the upper limit of skill of our forecast system during the period considered, as they include observed SST anomalies and external radiative forcing. Over all three land regions, the skill of the SST-restored simulations is generally higher and larger in extent than that of the hindcasts and twentieth century-RF simulations. Over North America and western Europe, the regions that show significantly enhanced skill over the twentieth century-RF simulations are larger than those of the hindcasts (Fig. 1a, d), indicating that considerably more skill may be achievable with better initialization techniques and improved models.

Skill in predicting North Atlantic SST on decadal timescales would imply, as described above, skill in initializing the Atlantic MOC. In the SST-restored simulations, the Atlantic MOC weakens from the

<sup>1</sup>Leibniz Institute of Marine Sciences, Düsternbrooker Weg 20, D-24105 Kiel, Germany. <sup>2</sup>Max Planck Institute for Meteorology, Bundesstraße 53, 20146 Hamburg, Germany.



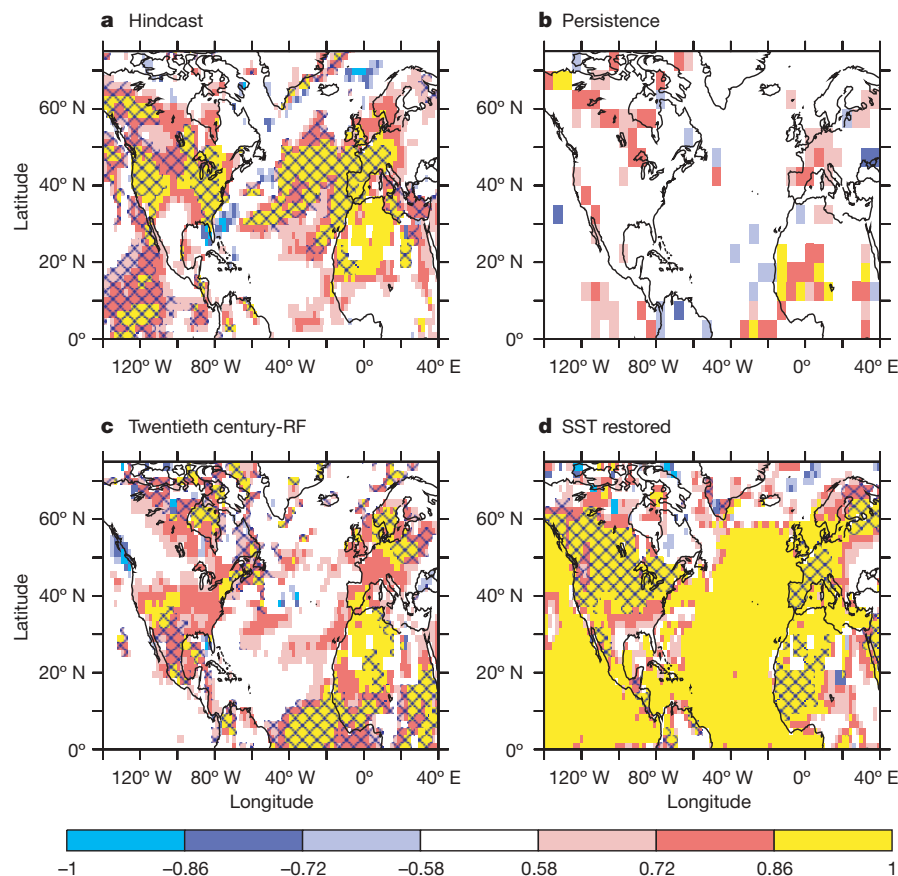
1950s till the mid-1960s, strengthens thereafter, peaking between the late-1980s to late-1990s, and subsequently weakens again (Fig. 2a). Although consistent with SST observations<sup>25</sup>, the veracity of these results is hard to assess, as direct MOC observations are insufficient<sup>8</sup> and large uncertainties exist among ocean model simulations<sup>21</sup> and ocean analyses. The small ensemble spread relative to the low frequency MOC fluctuations also supports the utility of the initialization scheme.

In contrast to measurements of the MOC, regular hydrographic observations in the Labrador Sea extend back past the 1950s. As density changes in the Labrador Sea are widely accepted to force MOC variations<sup>5,21,25</sup>, these observations provide an alternative method to assess MOC initialization. Simulated multidecadal fluctuations of wintertime Labrador Sea convection, weakest around 1970 and strongest in the early 1990s, broadly agree with observations of Labrador Sea Water thickness<sup>26</sup>, which is closely related to convection (Fig. 2b). Furthermore, simulated MOC variations closely follow Labrador Sea convection by several years (Fig. 2). Thus, these results provide evidence that observed MOC fluctuations are, to a certain degree, skilfully initialized.

On multidecadal timescales, Labrador Sea convection is initialized directly by SST relaxation. The latter includes the history of observed

atmospheric forcing, and in particular the North Atlantic Oscillation (NAO), which plays a key role in forcing multidecadal MOC variations<sup>20,21,25</sup>. Consistently, observed NAO variations lead simulated MOC changes by several years, but inconsistently, they are not well related to simulated Labrador Sea convection, except perhaps on multidecadal timescales (Fig. 2). Although the simulated NAO index does not correspond well with the observations, it strengthens during the simulated period, and thus probably also contributes to forcing MOC variations (Fig. 2c).

Initialized decadal fluctuations in the Atlantic MOC are predictable a decade in advance, with ensemble-spread small compared to the signal (Fig. 3a); hindcast skill is, however, largely due to capturing the long-term trend. MOC variations in the twentieth century-RF simulations are weak, and not surprisingly, unrelated to those in the initialization simulations. Model studies indicate that multidecadal Atlantic MOC variations force inter-hemispheric dipolar SST anomalies<sup>5,7,14,16,19</sup>. Observed variations of the latter are predicted a decade in advance by the hindcasts, but not by the twentieth century-RF simulations (Fig. 3b). Hindcast skill here is thus consistent with predicting Atlantic MOC variations. Consistently, North Atlantic (0–60° N) SST is better predicted by the hindcasts ( $r = 0.77$ ) than the twentieth century-RF simulations ( $r = 0.66$ ) (not shown). The

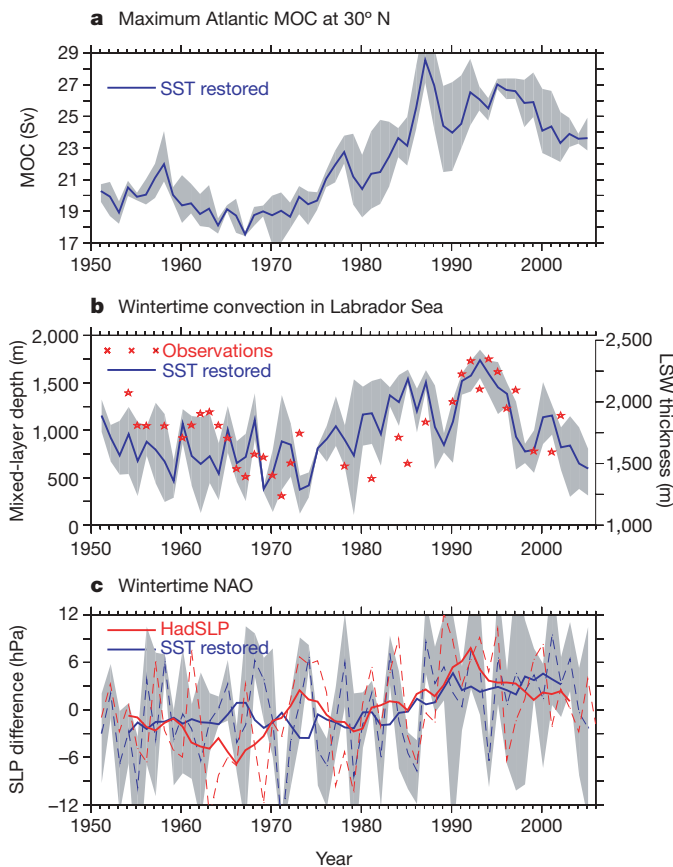


**Figure 1 | Correlation skill in predicting observed ten-year mean surface temperature anomalies a decade in advance relative to other more traditional approaches.** **a**, Skill of nine ten-year-long predictions, evenly distributed over the period 1955–2005, made with a climate model initialized using ocean (SST) observations and run with projected changes in radiative forcing. **b**, As in **a** but given by persistence. **c**, As in **a**, but not initialized using ocean observations and with radiative forcing following observations. **d**, As in **c**, but with model SST relaxed to observations between 60° S and 60° N (seen in near perfect correlations over the ocean). Correlations exceeding 0.58 are significant at the 5% level. Regions where initialization results in a significant enhancement or reduction in skill

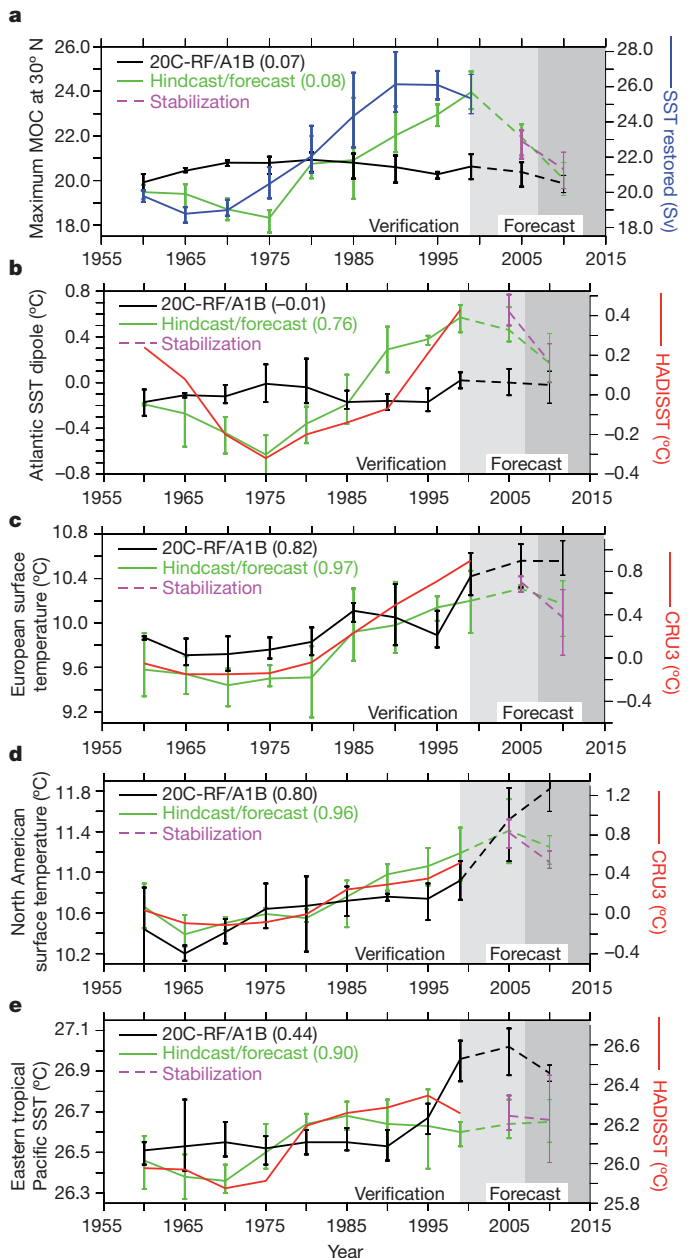
compared to radiative-forcing-only simulations are indicated by a blue cross-hatching in **a** and **c**, respectively. Land regions where restoring to observed SST anomalies provides a significant enhancement in skill relative to radiative-forcing-only simulations are indicated by blue cross-hatching in **d**. Correlations in **a** and **c** are field significant at close to the 0% level, while those in **b** pass the field significance test at the 1% level. Details of significance estimation are given in Methods. Correlations in **a**, **c** and **d** are computed from the ensemble mean of three simulations. SST observations are from HADISST<sup>27</sup>; land surface temperature observations are from CRUTEMP3<sup>28</sup>.

greater skill of the hindcasts, over the twentieth century-RF simulations, in predicting decadal North Atlantic, western European (Figs 1 and 3c) and North American (Figs 1 and 3d) surface-temperature variations is thus likely to be due to skill in predicting Atlantic MOC fluctuations.

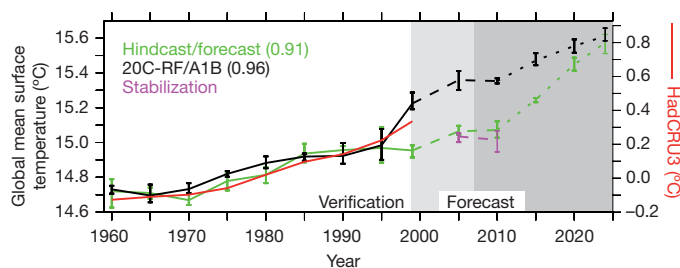
Enhanced skill over the twentieth century-RF simulations in predicting land surface temperature (Fig. 1a, c), especially over North America, may also arise from outside the Atlantic. In particular, tropical Pacific decadal SST variations are more skilfully predicted by the hindcasts, which correctly capture the 1970s climate shift, than by the twentieth century-RF simulations (Fig. 3e; Supplementary Information). Hindcast skill for global mean temperature (Fig. 4) is also high, but slightly less than the twentieth century-RF simulations. The largest difference occurs for the 1994–2004 predictions, with hindcast surface temperature cooler than observed and twentieth century-RF simulations warmer (Fig. 4). These differences stem from the predictions of internal tropical Pacific variability, with extreme warm conditions predicted by the twentieth century-RF simulations and near neutral conditions by the hindcasts;



**Figure 2 | SST restoring forces robust multidecadal fluctuations in the Atlantic Ocean meridional overturning circulation, Labrador Sea convection and the North Atlantic Oscillation.** **a**, Annual mean meridional overturning circulation (MOC) at 30° N from the SST-restored simulations. Grey shading indicates ensemble spread in all panels. **b**, Simulated wintertime (December–February) Labrador Sea (60–50° W, 55–65° N) mixed-layer depth and observed annual mean Labrador Sea Water (LSW) thickness<sup>26</sup>. The latter is defined between isopycnals  $\sigma_{1.5} = 34.72$ – $34.62$ , and is closely related to wintertime Labrador Sea convection. Simulated Labrador Sea convection precedes MOC variations at 30° N, with a maximum correlation (0.71) when the MOC lags by three years. **c**, Observed<sup>29</sup> and simulated ensemble mean wintertime North Atlantic Oscillation (NAO) indices. Solid lines indicate the seven-year running mean. The NAO index is defined as the sea-level-pressure (SLP) difference between Lisbon and Stykkisholmur. Observed NAO variations precede MOC variations at 30° N, with a maximum correlation (0.52) when the MOC lags by 3–4 years, but are not closely related to simulated Labrador Sea convection ( $r < 0.4$  at any lag).



**Figure 3 | Hindcast/forecast decadal means of selected time series, compared with observations and climate model projections including only radiative forcing changes but not initialized using ocean observations.** Model projections are twentieth century-RF followed by A1B scenario simulations ('20C-RF/A1B'). **a**, Maximum MOC strength at 30° N; **b**, Atlantic SST dipole index (60–10° W, 40–60° N minus 50–0° W, 40–60° S SST area averages), which is constructed to isolate MOC forced SST fluctuations from radiatively forced variations<sup>25</sup>; **c**, European land surface temperature (5° W–10° E, 35–60° N average); **d**, North American surface temperature (120–70° W, 30–50° N average); and **e**, eastern Tropical Pacific SST (150–90° W, 20° S–20° N average). **a–e**, Each point represents a ten-year centred mean; vertical bars indicate ensemble spread; verification and forecast periods are indicated (dark shading begins 2008, indicating the start of the true forecast period); two additional forecasts with radiative forcing stabilized at year 2000 levels are also shown. **b–e**, Correlation of both hindcasts and climate model projections with observations are given in brackets. Owing to insufficient observations, in **a** the correlation with the MOC from SST-restored simulation is given. Hindcast correlation skill is significantly greater at the 5% level than that of the standard climate projection in all cases (significance test described in Methods). Scale on right is used for the SST-restored simulation and observations; note the different ranges for right and left axes in **a** and **b**. SST are from HADISST<sup>27</sup>, land temperature from CRUTEMP<sup>38</sup> ('CRU3').



**Figure 4 | Hindcast/forecast decadal variations in global mean temperature, as compared with observations and standard climate model projections.** Model projections are twentieth century-RF followed by A1B scenario simulations ('20C-RF/A1B'); 'Stabilization' forecasts assume greenhouse gas concentrations fixed at year 2000 levels. Each point represents a ten-year centred mean; vertical bars indicate ensemble spread; verification and forecast periods are indicated (dark shading begins 2008, indicating the start of the true forecast period). Three additional decadal means (joined by a dotted line) show the evolution of the initialized and un-initialized 2005 predictions extended till 2030. Correlation of both hindcasts and climate model projections with observations are given in brackets. Correlation of the twentieth century-RF simulation with observations is greater than that of the hindcasts, but only marginally at the 5% significance level. Observed global mean temperature anomalies are from HadCRU3<sup>28</sup>.

weak warm conditions were observed (Fig. 3e; Supplementary Information).

We now consider two forecasts, started in November 2000 and November 2005. The MOC is predicted to weaken almost to its 1950–2005 mean over the next decade (Fig. 3a), leading to a weakening of Atlantic SST hemispheric difference towards zero (Fig. 3b). North Atlantic (not shown), western European (Fig. 3c) and North American (Fig. 3d) surface temperatures cool towards 1994–2004 levels. In contrast, in the un-initialized (twentieth century-RF) predictions the MOC slightly weakens, the hemispheric SST difference is unchanged, and warming of surface temperatures over the latter three regions continues (Fig. 3a–d). Eastern tropical Pacific SST is forecasted to remain almost unchanged, but 0.3 K cooler than the un-initialized predictions (Fig. 3e). The differences in predicted North Atlantic and tropical Pacific variability lead to a large difference in the global mean temperature prediction: the initialized prediction indicates a slight cooling relative to 1994–2004 levels, while the anthropogenic-forcing-only simulation suggests a near 0.3 K rise (Fig. 4). In the long-term both projections agree with each other, as is found by extending the 2005 prediction till 2030 (Fig. 4). Internal decadal fluctuations were also found to offset anthropogenic global warming in a previous study<sup>19</sup>, but the offset was much less pronounced and associated primarily with changes in the tropical Pacific.

To investigate the sensitivity of the predictions to greenhouse gas forcing only, the two forecasts were repeated assuming that greenhouse gases were stabilized at year 2000 values. The predictions for the MOC and surface temperature remain basically unchanged. Thus, in the near future, natural decadal variability in the Atlantic and Pacific may not only override the regional effects of global warming, but temporarily weaken it. Thus, a joint initial/boundary value problem has to be considered when forecasting North Atlantic sector and global climate variability for the coming decades.

The results presented here are promising in light of existing model biases. Experience in numerical weather prediction and seasonal forecasting has shown that skill can be considerably improved by reducing model systematic error and by more accurate forecast initialization. Thus, useful decadal predictions may be in reach.

## METHODS SUMMARY

**Simulations.** Results are based on three sets of simulations with the ECHAM5/MPI-OM coupled general circulation model<sup>24</sup> (IPCC version). (1) Three integrations (twentieth century-RF/A1B), from 1860–2015, started from different points of a control simulation. Before 2000, radiative forcing follows

observations (greenhouse gas and sulphate aerosol concentrations, solar cycle variations, and major volcanic eruptions), and after 2000, it follows the IPCC A1B scenario. Predictability due to radiative forcing is estimated from these simulations. (2) Three coupled integrations (SST-restored), from 1950–2005, initialized from the twentieth century-RF simulations and with identical radiative forcing, but with relaxation towards SST constructed from the coupled model climatology with observed SST anomalies superimposed. The relaxation constant is strong ( $0.25 \text{ d}^{-1}$ ) between  $30^\circ \text{ S}$  and  $30^\circ \text{ N}$ , and decreases linearly to zero between  $30^\circ$  and  $60^\circ \text{ S}$  and between  $30^\circ$  and  $60^\circ \text{ N}$ ; poleward of  $60^\circ \text{ S}$  and  $60^\circ \text{ N}$  the model is fully coupled. (3) Nine hindcasts and two forecasts (three-member, ten-years long), initialized from the SST-restored simulations during the period 1955–2005. Radiative forcing is as in the twentieth century-RF simulations, except for solar cycle variations, which are repeated from the previous 11 years, and major volcanic eruptions. The latter that occurred during a hindcast are not included, and the impact of any that occurred before the hindcast is damped away with an e-folding time of one year.

**Significance tests.** For nine forecasts we assume seven degrees of freedom (d.f.). As only positive correlations indicate skill, the significance level for the forecast is determined using a one-sided *t*-test. A correlation is determined to be significantly larger or smaller than another if the Fisher-Z transformed values pass a one-sided *t*-test for differences in means (with 12 d.f.). The Fisher-Z distribution of correlations of nine independent and normally distributed pairs of points is close to normal. Field significance is estimated on a  $15^\circ \times 15^\circ$  grid to minimize effects of spatial correlation.

Received 25 June 2007; accepted 14 March 2008.

1. Goldenberg, S. B., Landsea, C. W., Mestas-Nuñez, A. M. & Gray, W. M. The recent increase in Atlantic hurricane activity: Causes and implications. *Science* **293**, 474–479 (2001).
2. Enfield, D. B., Mestas-Nuñez, A. M. & Trimble, P. J. The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U. S. *Geophys. Res. Lett.* **28**, 2077–2080 (2001).
3. Sutton, R. T. & Hodson, D. L. R. Atlantic Ocean forcing of North American and European summer climate. *Science* **309**, 115–118 (2005).
4. Folland, C. K., Palmer, T. N. & Parker, D. E. Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature* **320**, 602–607 (1986).
5. Griffies, S. M. & Bryan, K. Predictability of North Atlantic multidecadal climate variability. *Science* **275**, 181–184 (1997).
6. Boer, G. A study of atmosphere-ocean predictability on long time scales. *Clim. Dyn.* **16**, 469–472 (2000).
7. Collins, M. *et al.* Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study. *J. Clim.* **19**, 1195–1203 (2006).
8. Cunningham, S. A. *et al.* Temporal variability of the Atlantic meridional overturning circulation at  $26.5^\circ \text{ N}$ . *Science* **317**, 935–938 (2007).
9. Smith, D. M. *et al.* Improved surface temperature prediction for the coming decade from a global climate model. *Science* **317**, 796–799 (2007).
10. Solomon, S. *et al.* *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, Cambridge, UK, 2007).
11. Bjerknes, J. Atlantic air-sea interaction. *Adv. Geophys.* **10**, 1–82 (1964).
12. Kushnir, Y. Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.* **7**, 141–157 (1994).
13. Schlesinger, M. E. & Ramankutty, N. An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).
14. Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M. & Mann, M. E. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024233 (2005).
15. Mann, M. E., Bradley, R. S. & Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
16. Delworth, T., Manabe, S. & Stouffer, R. J. Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model. *J. Clim.* **6**, 1993–2011 (1993).
17. Vellinga, M. & Wu, P. Low-latitude freshwater influence on centennial variability of the Atlantic thermohaline circulation. *J. Clim.* **17**, 4498–4511 (2004).
18. Jungclauss, J. H., Haak, H., Latif, M. & Mikolajewicz, U. Arctic North Atlantic interactions and multidecadal variability of the meridional overturning circulation. *J. Clim.* **18**, 4013–4031 (2005).
19. Latif, M., Collins, M., Pohlmann, H. & Keenlyside, N. A review of predictability studies of the Atlantic sector climate on decadal time scales. *J. Clim.* **19**, 5971–5987 (2006).
20. Eden, C. & Jung, T. North Atlantic interdecadal variability: Oceanic response to the North Atlantic oscillation (1865–1997). *J. Clim.* **14**, 676–691 (2001).
21. de Coëtlogon, G. *et al.* Gulf Stream variability in five oceanic general circulation models. *J. Phys. Oceanogr.* **36**, 2119–2135 (2006).
22. Keenlyside, N., Latif, M., Botzet, M., Jungclauss, J. & Schulzweida, U. A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature. *Tellus A* **57**, 340–356 (2005).
23. Lee, T. C. K., Zwiers, F., Zhang, X. & Tsao, M. Evidence of decadal climate prediction skill resulting from changes in anthropogenic forcing. *J. Clim.* **19**, 5305–5318 (2006).



24. Jungclauss, J. H. *et al.* Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *J. Clim.* **19**, 3952–3972 (2006).
25. Latif, M. *et al.* Is the thermohaline circulation changing? *J. Clim.* **19**, 4631–4637 (2006).
26. Curry, R. G., McCartney, M. S. & Joyce, T. M. Oceanic transport of subpolar climate signals to mid-depth subtropical waters. *Nature* **391**, 575–577 (1998).
27. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407, doi:10.1029/2002JD002670 (2003).
28. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548 (2006).
29. Allan, R. & Ansell, T. A new globally-complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2004. *J. Clim.* **19**, 5816–5842 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Computing resources at the European Centre for Medium Range Weather Forecasting, the Deutsches Klimarechenzentrum and the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen are acknowledged. Deutscher Wetterdienst provided significant CPU allocations. The work was supported by the European Union ENSEMBLES and German BMBF NORDATLANTIK projects. We thank M. Esch for performing the three twentieth century-RF simulations, and F. Nevoigt for help with art work.

**Author Contributions** N.S.K. and M.L. designed experiments. N.S.K. performed experiments and analysis. M.L. and J.J. contributed to analysis. L.K., J.J. and E.R. provided the model and support. N.S.K. and M.L. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to N.S.K. ([nkeenlyside@ifm-geomar](mailto:nkeenlyside@ifm-geomar)).

# Origin of a 'Southern Hemisphere' geochemical signature in the Arctic upper mantle

Steven L. Goldstein<sup>1,2</sup>, Gad Soffer<sup>1,2,†</sup>, Charles H. Langmuir<sup>3</sup>, Kerstin A. Lehnert<sup>1</sup>, David W. Graham<sup>4</sup> & Peter J. Michael<sup>5</sup>

The Gakkel ridge, which extends under the Arctic ice cap for ~1,800 km, is the slowest spreading ocean ridge on Earth. Its spreading created the Eurasian basin, which is isolated from the rest of the oceanic mantle by North America, Eurasia and the Lomonosov ridge. The Gakkel ridge thus provides unique opportunities to investigate the composition of the sub-Arctic mantle and mantle heterogeneity and melting at the lower limits of sea-floor spreading. The first results of the 2001 Arctic Mid-Ocean Ridge Expedition (ref. 1) divided the Gakkel ridge into three tectonic segments, composed of robust western and eastern volcanic zones separated by a 'sparsely magmatic zone'. On the basis of Sr–Nd–Pb isotope ratios and trace elements in basalts from the spreading axis, we show that the sparsely magmatic zone contains an abrupt mantle compositional boundary. Basalts to the west of the boundary display affinities to the Southern Hemisphere 'Dupal' isotopic province<sup>2</sup>, whereas those to the east—closest to the Eurasian continent and where the spreading rate is slowest—display affinities to 'Northern Hemisphere' ridges. The western zone is the only known spreading ridge outside the Southern Hemisphere that samples a significant upper-mantle region with Dupal-like characteristics. Although the cause of Dupal mantle has been long debated, we show that the source of this signature beneath the western Gakkel ridge was subcontinental lithospheric mantle that delaminated and became integrated into the convecting Arctic asthenosphere. This occurred as North Atlantic mantle propagated north into the Arctic during the separation of Svalbard and Greenland.

The Gakkel ridge is the global ultraslow endmember among ocean ridges, with spreading rates decreasing from ~15 mm yr<sup>-1</sup> (full rate) at the Lena trough near Greenland to ~6 mm yr<sup>-1</sup> at the Siberian margin. It contains no significant transform offsets, which on other ridges often form mantle flow and mantle compositional boundaries. It thus provides a natural laboratory in which to study source heterogeneity and melting dynamics at ultraslow spreading rates, and where potential complications from tectonic offsets are absent. Here we report isotope and trace element data on basaltic glasses that document the Arctic mantle composition and processes contributing to its enrichment.

Prior to the 2001 Arctic Mid-Ocean Ridge Expedition, data on Gakkel lavas were restricted to three samples of basaltic cobbles and glass shards recovered on two cruises<sup>3</sup>. The Arctic Mid-Ocean Ridge Expedition collected samples from the western ~850 km of the ridge between longitudes 7° W and 86° E at over 200 stations. Michael *et al.*<sup>1</sup> divided the studied segments into three 'magmato-tectonic zones': the ~200-km-long, shallow, magmatically robust western volcanic zone (WVZ), characterized by normal ocean-ridge-rift

magmatism; the ~300-km-long, deep, magma-starved sparsely magmatic zone (SMZ) with peridotite outcropping along the ridge axis; and the ~400-km-long, magmatically robust eastern volcanic zone (EVZ), where axial magmatism is mainly associated with central volcanoes. On the basis of cruise chemical data, Michael *et al.*<sup>1</sup> inferred compositionally distinct mantle beneath the WVZ and the EVZ, with marked barium enrichments in WVZ lavas. Dick *et al.*<sup>4</sup> compared the broad-scale tectonic features of the Gakkel and southwest Indian ridges and posited the existence of a new class of ultraslow-spreading ocean ridge, where 'amagmatic' segments like the SMZ are a consequence of spreading rates less than ~12 mm yr<sup>-1</sup>, and are thus independent of mantle composition.

Our data confirm fundamental differences in the upper-mantle sources of WVZ and EMZ lavas (locations and data tables can be found in the Supplementary Information). The fields formed by Nd–Sr isotopes of EVZ and WVZ lavas show no overlap (Fig. 1a), and samples from the western and eastern parts of the SMZ show affinities to the adjacent EVZ and WVZ. EVZ lavas have higher <sup>143</sup>Nd/<sup>144</sup>Nd values than do WVZ lavas and lower <sup>87</sup>Sr/<sup>86</sup>Sr values than does any mid-ocean-ridge basalt (MORB) north of Iceland.

The contrasting WVZ and EVZ mantle compositions are shown in Fig. 1b in terms of  $\Delta 8/4$  (ref. 2), which reflects deviations from typical compositions of Pacific and Atlantic MORB on <sup>208</sup>Pb/<sup>204</sup>Pb–<sup>206</sup>Pb/<sup>204</sup>Pb diagrams. The disparities also show in along-axis plots of the other radiogenic isotopes (see, for example, Supplementary Fig. 1a). West of the boundary virtually all Gakkel lavas have  $\Delta 8/4 > 30$ ; this was shown by Hart<sup>2</sup> to be a primary feature of the Southern Hemisphere 'Dupal isotope anomaly' in Indian and some South Atlantic MORB. The Dupal-like character of the WVZ is further illustrated by comparison with global MORB: <sup>208</sup>Pb/<sup>204</sup>Pb–<sup>206</sup>Pb/<sup>204</sup>Pb data for WVZ basalts lie in the field of Indian MORB (Fig. 2a; Supplementary Fig. 2a). WVZ lavas also show a near-vertical trend in a <sup>87</sup>Sr/<sup>86</sup>Sr–<sup>206</sup>Pb/<sup>204</sup>Pb diagram, reflecting a large range of <sup>87</sup>Sr/<sup>86</sup>Sr values, typical of Indian MORB, whereas EVZ lavas show a shallow trend similar to that of Atlantic–Pacific basalts (Fig. 2b). Additional evidence of the WVZ's Dupal-like character is given in Supplementary Information (Supplementary Fig. 3).

Gakkel WVZ lavas, however, differ from Southern Hemisphere Dupal basalts, in that they have low  $\Delta 7/4$  values<sup>2</sup>; their data lie near Atlantic–Pacific basalts on this diagram (Supplementary Fig. 2b). The <sup>207</sup>Pb/<sup>204</sup>Pb–<sup>206</sup>Pb/<sup>204</sup>Pb relationships are temporal rather than a result of the formation process, because they depend primarily on the timing of major U–Pb fractionation, whereas  $\Delta 8/4$  reflects the chemical composition (Th/U) and the formation age. Oceanic basalts with enriched isotopic signatures (high Sr and low Nd and Hf isotope ratios, and high  $\Delta 8/4$  values) are associated with both high and low

<sup>1</sup>Lamont-Doherty Earth Observatory of Columbia University, <sup>2</sup>Department of Earth and Environmental Sciences, Columbia University, 61 Route 9W, Palisades, New York 10964, USA.

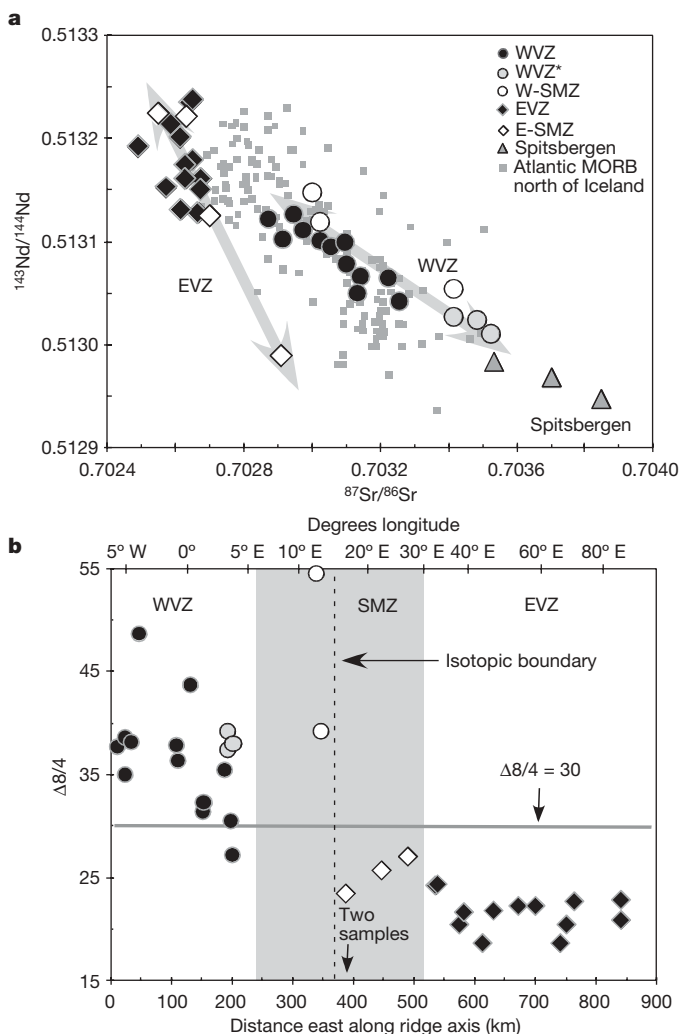
<sup>3</sup>Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. <sup>4</sup>College of Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. <sup>5</sup>Department of Geosciences, University of Tulsa, 600 South College Avenue Tulsa, Oklahoma 74104, USA. †Present address: Harvard Business School, Soldiers Field, Boston, Massachusetts 02163, USA.

$\Delta 7/4$  values (ref. 5), reflecting old and young formation ages, respectively. The Gakkel data thus indicate that the formation of the Arctic Dupal-like source was more recent than that of the Southern Hemisphere Dupal basalts. Although the formation ages differ, chemical characteristics are similar: like Indian MORB, the high  $\Delta 8/4$  and  $^{87}\text{Sr}/^{86}\text{Sr}$  values reflect high time-integrated Th/U and Rb/Sr values for the Pb and Sr in the lavas, indicating the distinctive chemical compositions of reservoirs responsible for Dupal features.

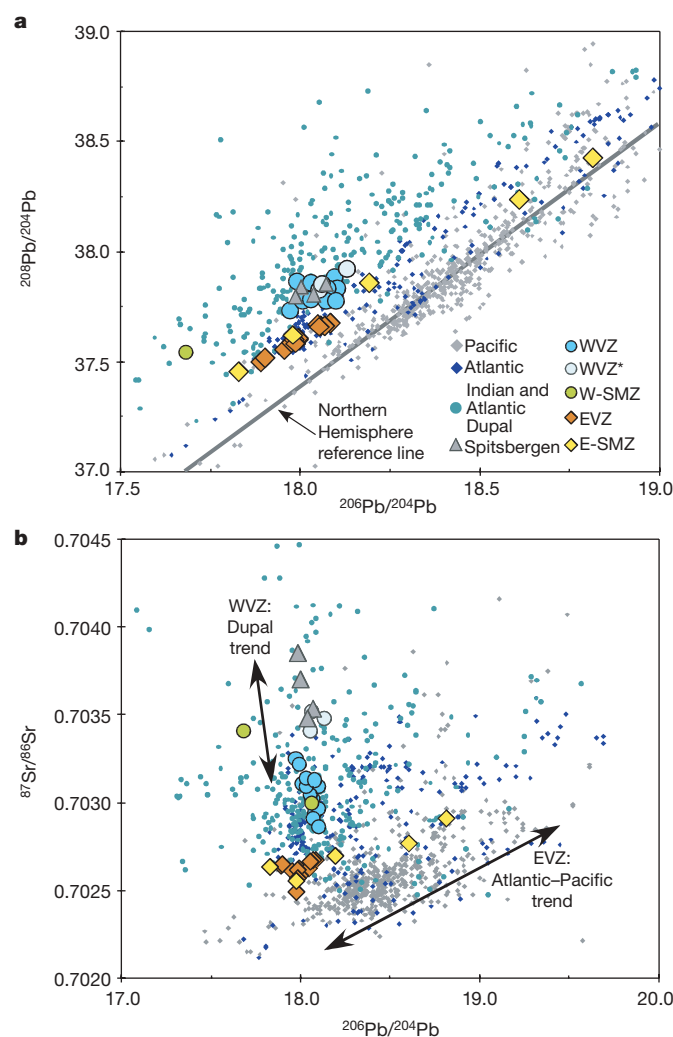
A striking feature of the Gakkel ridge is the abruptness of the isotopic boundary and its location in the middle of the SMZ (Fig. 1b; Supplementary Fig. 1a). The discontinuity is reminiscent of the Australian–Antarctic discordance of the southeast Indian ridge, which demarcates the boundary between the Indian Dupal and Pacific mantle provinces<sup>5</sup>. Common features include great

bathymetric depths and low magma production, as reflected by the basaltic crust being thin in comparison with that of adjoining segments. These indicate close relationships between seafloor structure, bathymetry and mantle convective processes. However, the Gakkel SMZ, unlike the Australian–Antarctic discordance, is not associated with any significant ridge offset. Along a continuously spreading ridge, lavas with isotopic affinities to the WVZ or EVZ were collected within 40 km of each other at longitudes 12.7° E and 16.0° E. The Gakkel SMZ data thus indicate a sharp geochemical boundary in the absence of any tectonic boundary. The Knipovich ridge to the south-west of the Gakkel ridge is also isotopically distinct from the WVZ (see, for example, Supplementary Figs 1b, 2a and 3a), but these ridges are separated by a 600 km offset. The marked change in the magmato-tectonic character at the Gakkel SMZ is therefore not simply a tectonic reflection of ultraslow seafloor spreading, as proposed by Dick *et al.*<sup>4</sup>, but reflects a mantle compositional and convective boundary.

Comparison with global MORB (Fig. 2) shows that the Pb–Sr–Nd isotope variations in this small Arctic region mimic the global upper-mantle variability. The Gakkel ridge is far from the Indian Ocean. Although enriched MORBs are found along North Atlantic ridges

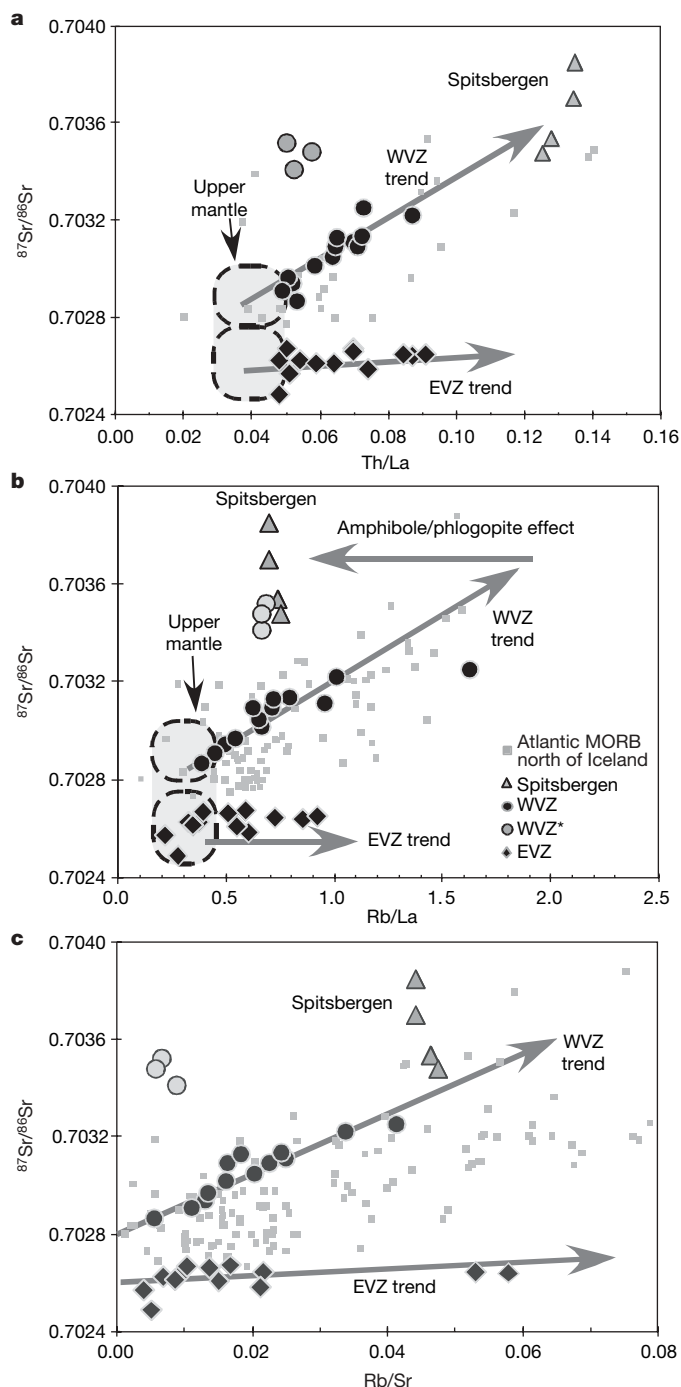


**Figure 1 | Distinct western and eastern geochemical provinces in Gakkel ridge basalts.** **a**,  $^{143}\text{Nd}/^{144}\text{Nd}$  versus  $^{87}\text{Sr}/^{86}\text{Sr}$ . Data fields for EVZ and WVZ lavas do not overlap. WVZ lava data points overlap with those for Atlantic MORBs north of Iceland and display a trend towards those for Spitsbergen Quaternary-period alkaline basalts. SMZ lavas show affinities to those in the adjacent WVZ and EVZ. W-SMZ and E-SMZ refer to lavas in the SMZ having isotopic affinities to the WVZ and EVZ, respectively. **b**,  $\Delta 8/4$  (representing deviations from the Northern Hemisphere reference line<sup>2</sup>) versus distance from the western end of the Gakkel ridge near Greenland. (The key is the same as that in **a**.) This shows the distinct WVZ and EVZ compositions and the position and abruptness of the isotopic boundary. Three samples, labelled WVZ\*, from near the WVZ–SMZ boundary have unusual chemistry (low  $\text{SiO}_2$  and incompatible element abundances, and positive Eu and Sr anomalies on trace element diagrams) and follow the WVZ isotope trend (**a**) but not the WVZ chemical trends (Fig. 3).



**Figure 2 | Gakkel basalts as a microcosm of global MORBs.** **a**,  $^{208}\text{Pb}/^{204}\text{Pb}$  versus  $^{206}\text{Pb}/^{204}\text{Pb}$ . WVZ and Spitsbergen lavas show the characteristic Dupal MORB offset to high  $^{208}\text{Pb}/^{204}\text{Pb}$ . Global MORB data are from the Petrological Database of the Ocean Floor ([www.petdb.org](http://www.petdb.org)). **b**,  $^{87}\text{Sr}/^{86}\text{Sr}$  versus  $^{206}\text{Pb}/^{204}\text{Pb}$ . (The key is the same as that in **a**.) EVZ lavas display a trend similar to that of Atlantic–Pacific MORBs, whereas WVZ MORBs display a near-vertical trend similar to Indian and South Atlantic MORBs showing the Dupal anomaly. The WVZ data points display a trend towards those for Spitsbergen lavas.





**Figure 3 | Gakkel ridge and Spitsbergen basalts.** **a**,  $^{87}\text{Sr}/^{86}\text{Sr}$  versus Th/La. (The key is the same as that in **b**.) WVZ and EVZ lavas form distinct trends. WVZ data points display a trend towards those for Spitsbergen, indicating that the Spitsbergen magma source is the enriched WVZ endmember. **b**,  $^{87}\text{Sr}/^{86}\text{Sr}$  versus Rb/La. WVZ and EVZ lavas display distinct trends. Spitsbergen lava data points lie away from the WVZ trend, with low Rb/La values (and Ba/La values, not shown), which is consistent with partial melting in the presence of residual amphibole and/or phlogopite, and indicates a SCLM source. **c**,  $^{87}\text{Sr}/^{86}\text{Sr}$  versus Rb/Sr. (The key is the same as that in **b**.) The Sr isotope ratios are distinct for the WVZ and EVZ even when Rb/Sr = 0. In **a** and **b**, for low Th/La and Rb/La values corresponding to depleted mantle composition estimates<sup>32,33</sup> (labelled 'upper mantle'), the WVZ and EVZ Sr isotope ratios also are different. In all three panels, 'depleted' WVZ compositions are consistent with Atlantic MORBs from north of Iceland, but EVZ compositions are distinct. MORB data are taken from the Petrological Database of the Ocean Floor.

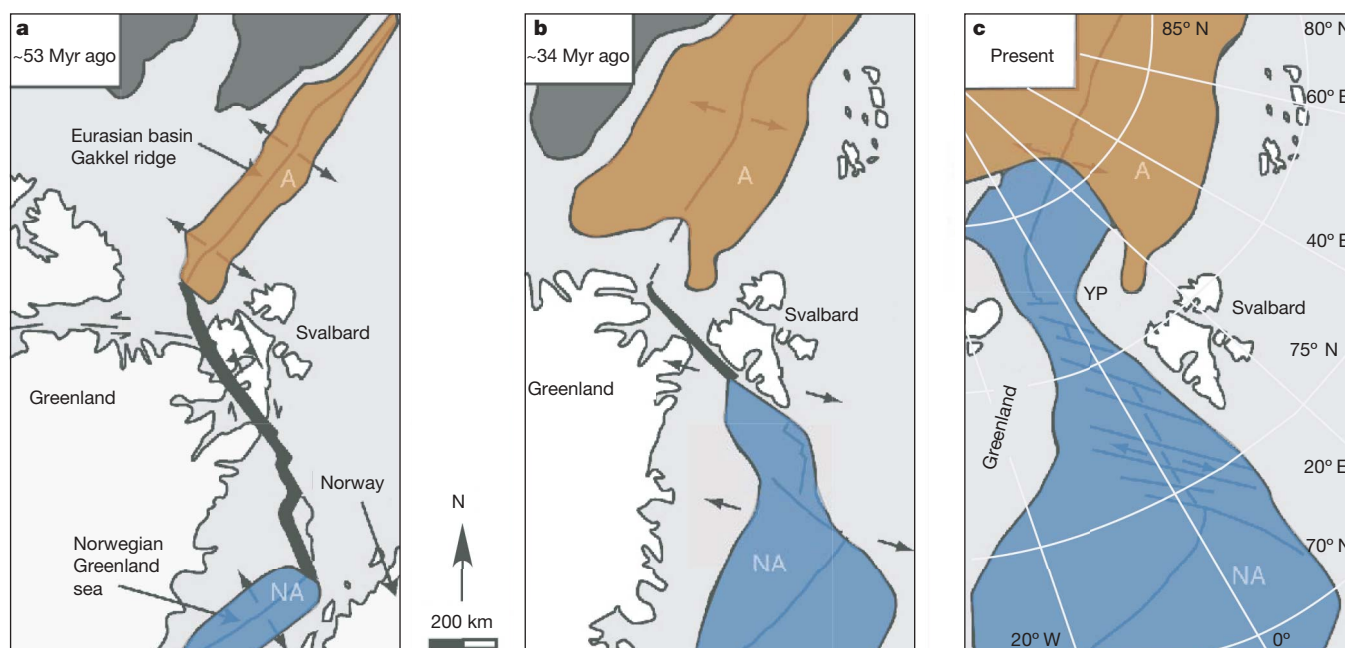
even far from hotspots<sup>6,7</sup>, the WVZ is unique among North Atlantic MORB, with Dupal characteristics dominating along ~350 km of the spreading axis. Their occurrence in Arctic MORB shows that Dupal characteristics are not simply an attribute of the Indian–South Atlantic oceanic mantle, but reflect processes that are more general.

The small geographical scale and well-understood tectonic history of the Eurasian basin affords opportunities to explore how the Dupal signature is created. Quaternary alkaline basalts ranging from hawaiites to nephelinitic basanites in northwestern Spitsbergen have been attributed to low-degree melting of subcontinental lithospheric mantle (SCLM)<sup>8–10</sup>, and form an enriched endmember for WVZ lavas in isotope plots (Figs 1a and 2). This relationship also holds for trace elements that are highly incompatible for melting of both anhydrous and metasomatized mantle mineral assemblages, where abundance ratios in lavas are almost the same as in the mantle source (shown for  $^{87}\text{Sr}/^{86}\text{Sr}$  versus Th/La and U/La in Fig. 3a and Supplementary Fig. 4a, respectively). The trace elements and isotopes together strongly indicate that the component causing the Dupal-like characteristics is also the source of Spitsbergen alkaline basalts.

Insight into the mineral assemblage of the Spitsbergen magma source can be gleaned from the comparison of elements more compatible in amphibole or phlogopite (such as Rb, Ba and K) with other elements unaffected by amphibole or phlogopite (such as Th, U and La). In plots of  $^{87}\text{Sr}/^{86}\text{Sr}$  versus Rb/La (Fig. 3b) and Ba/Th (Supplementary Fig. 4b), Spitsbergen lavas deviate from the WVZ trend at low values of Rb/La and Ba/Th, respectively; however, they are not offset in plots involving Th, U and La. Such chemical behaviour indicates the presence of amphibole or phlogopite among the residual minerals<sup>11,12</sup>. Spitsbergen lavas include olivine, titanite and plagioclase, but no amphibole or phlogopite<sup>8</sup>. Because these phases are unstable at asthenospheric temperatures<sup>12,13</sup>, the data require an SCLM source for Spitsbergen lavas. Subsequent asthenospheric melting beneath the Gakkel ridge would then release the Rb and Ba, leading to the relative enrichments of these elements in the lavas of the WVZ.

The data offer insight into the long-term history of the Gakkel asthenosphere. A Rb–Sr 'pseudo-isochron' plot (Fig. 3c) shows that the WVZ mantle least contaminated by SCLM is distinct from the EVZ mantle, because their respective  $^{87}\text{Sr}/^{86}\text{Sr}$  values differ when Rb/Sr = 0. This WVZ–EVZ offset is further confirmed by comparing isotopes and trace elements (Fig. 3a, b and Supplementary Fig. 4). The 'least contaminated' WVZ lavas are similar to Atlantic MORB north of Iceland, in terms of Nd–Sr isotopes and combined isotopes and trace elements (Figs 1 and 3 and Supplementary Fig. 4), whereas the EVZ basalts are distinct from both groups. These observations suggest that WVZ isotopic relationships can be explained by the addition of Spitsbergen SCLM to the North Atlantic asthenosphere, and that the WVZ and EVZ are distinct mantle provinces. The Gakkel SMZ thus bounds distinct upper-mantle regions that have experienced long-term convective isolation.

The introduction of the Svalbard SCLM to the WVZ asthenosphere, the leakage of North Atlantic mantle into the Arctic, and the formation of the boundary can be understood in the context of Arctic tectonic history (Fig. 4). The Eurasian and Atlantic oceans were separated by continent when seafloor spreading began in the Arctic and North Atlantic 55–60 million years (Myr) ago<sup>14</sup>. During the early Oligocene epoch ~34 Myr ago, Greenland and Norway–Svalbard began to separate, accompanied by northwards propagation of North Atlantic ridges, with linkage of the Arctic and North Atlantic occurring ~10 Myr ago. Our data suggest that North Atlantic mantle flowed into the Arctic, accompanied by delamination and dispersal of the Svalbard SCLM. Whether this was simply a consequence of rifting or was facilitated by thermal erosion caused by an Arctic hotspot is beyond the scope of this work (a hotspot has been postulated to be responsible for the Yermak plateau (Fig. 4c), but its existence is disputed<sup>15</sup>). In any case, the Gakkel SMZ appears to be the surface



**Figure 4 | Cenozoic-era evolution of the Arctic upper mantle.**

**a**, Approximately 53 Myr ago, the nascent Eurasian basin, underlain by 'Arctic' upper mantle (labelled 'A') as sampled today in the EVZ, was separated from the North Atlantic ridges (labelled 'NA') by a continent. **b**, During the Cenozoic, the Eurasian basin continued to widen, and the North Atlantic ridges propagated north, with Greenland and Svalbard

separating ~10 Myr ago. **c**, Present-day situation based on the Gakkel data, with the boundary between the North Atlantic and Arctic upper-mantle provinces well within the Eurasian basin, in the middle of the SMZ. 'YP' labels the Yermak plateau, near the WVZ. Images modified from Blythe and Kleinspehn<sup>14</sup>.

manifestation of a physical boundary within the mantle where the linkage occurred.

The origin of the Southern Hemisphere Dupal signature is an ongoing debate in mantle dynamics, and a recent summary highlights inadequacies of postulated models<sup>16</sup>. Explanations<sup>16–28</sup> include upwelling of a thermal anomaly from the core–mantle boundary; contamination of sub-Indian mantle by the Kerguelen plume; mantle modification by subduction surrounding Gondwana; ancient subduction of altered oceanic crust and sediment; recycled lower-continental or arc crust; and detachment of SCLM during continental break-up. Hybrid scenarios invoke both subduction and lithosphere delamination<sup>29–31</sup>. It has been suggested that SCLM delamination would impart 'too low' an osmium isotope signature to MORB<sup>26,27</sup>, but this depends on the age of the recycled SCLM. The combined Gakkel and Spitsbergen data and the well-constrained Arctic tectonic history provide relationships where the geochemistry and tectonics are in remarkable accord.

In summary, the Gakkel ridge is an ultraslow-spreading ridge surrounded by continental lithosphere, with clear evidence of regional SCLM contributing a Dupal geochemical signature to the asthenospheric mantle. Unlike in other regions on Earth, we have the strong evidence of Spitsbergen volcanics, with the requisite isotopic compositions to be the Gakkel endmember, and a source mineralogy definitively indicating SCLM melting. The process documented in the Arctic is chemically and geologically feasible and probably plays an important role in generating chemical heterogeneity along the mid-ocean ridge system. Dupal signatures globally may have multiple sources; nevertheless, the migrating continents that have traversed the Indian Ocean basin, where Dupal characteristics occur on a large scale, would generate substantial opportunities for analogous SCLM contributions to the sub-Indian Ocean mantle.

## METHODS SUMMARY

All analyses were made on hand-picked basaltic glasses. Major elements were measured on a Cameca electron microprobe at the American Museum of Natural History, operated as a joint facility with the Lamont-Doherty Earth Observatory

of Columbia University (LDEO). Trace elements were measured using a Plasma Quad 2 inductively coupled plasma mass spectrometer at LDEO or a Thermo Electron X-series inductively coupled plasma mass spectrometer at Harvard. Pb, Nd and Sr isotope ratios were measured using a VG Sector 54 multicollector thermal ionization mass spectrometer at LDEO.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 15 January; accepted 14 March 2008.

1. Michael, P. J. *et al.* Magmatic and amagmatic seafloor generation at the ultraslow-spreading Gakkel ridge, Arctic Ocean. *Nature* **423**, 956–961 (2003).
2. Hart, S. R. A large-scale isotopic anomaly in the Southern Hemisphere mantle. *Nature* **309**, 753–757 (1984).
3. Mühe, R., Bohrmann, H., Garbe-Schönberg, D. & Kassens, H. E-MORB glasses from the Gakkel Ridge (Arctic Ocean) at 87°N: evidence for the Earth's most northerly volcanic activity. *Earth Planet. Sci. Lett.* **152**, 1–9 (1997).
4. Dick, H. J. B., Lin, J. & Schouten, H. An ultraslow-spreading class of ocean ridge. *Nature* **426**, 405–412 (2003).
5. Klein, E. M., Langmuir, C. H., Zindler, A., Staudigal, H. & Hamelin, B. Isotope evidence of a mantle convection boundary at the Australian–Antarctic Discordance. *Nature* **333**, 623–629 (1988).
6. Dosso, L. *et al.* The age and distribution of mantle heterogeneity along the Mid-Atlantic Ridge (31–41°N). *Earth Planet. Sci. Lett.* **170**, 269–286 (1999).
7. Shirey, S. B., Bender, J. F. & Langmuir, C. H. Three-component isotopic heterogeneity near the Oceanographer transform, Mid-Atlantic Ridge. *Nature* **325**, 217–223 (1987).
8. Skjelkvale, B. L., Amundsen, H. E. F., O'Reilly, S. Y., Griffin, W. L. & Gjelsvik, T. A primitive alkali basaltic stratovolcano and associated eruptive centers, Northwestern Spitsbergen: Volcanology and tectonic significance. *J. Volcanol. Geotherm. Res.* **37**, 1–19 (1989).
9. Ionov, D. A., Bodinier, J. L., Mukasa, S. B. & Zanetti, A. Mechanisms and sources of mantle metasomatism: Major and trace element compositions of peridotite xenoliths from Spitsbergen in the context of numerical modeling. *J. Petrol.* **43**, 2219–2259 (2002).
10. Ionov, D. A., Mukasa, S. B. & Bodinier, J. L. Sr–Nd–Pb isotopic compositions of peridotite xenoliths from Spitsbergen: Numerical modelling indicates Sr–Nd decoupling in the mantle by melt percolation metasomatism. *J. Petrol.* **43**, 2261–2278 (2002).
11. Clague, D. A. & Frey, F. A. Petrology and trace element chemistry of the Honolulu volcanics, Oahu: Implication for the oceanic mantle below Hawaii. *J. Petrol.* **23**, 447–504 (1982).

12. Class, C. & Goldstein, S. L. Plume-lithosphere interactions in the ocean basins: constraints from the source mineralogy. *Earth Planet. Sci. Lett.* **150**, 245–260 (1997).
13. Niida, K. & Green, D. H. Stability and chemical composition of pargasitic amphibole in MORB pyroxene under upper mantle conditions. *Contrib. Mineral. Petrol.* **135**, 18–40 (1999).
14. Blythe, A. E. & Kleinspehn, K. L. Tectonically versus climatically driven exhumation of the Eurasian plate margin, Svalbard: Fission track analyses. *Tectonics* **17**, 621–639 (1998).
15. Ritzmann, O. & Jokat, W. Crustal structure of northwestern Svalbard and the adjacent Yermak Plateau: evidence for Oligocene detachment tectonics and non-volcanic breakup. *Geophys. J. Int.* **152**, 139–159 (2003).
16. Zhang, S. Q. *et al.* Evidence for a widespread Tethyan upper mantle with Indian-Ocean-type isotopic characteristics. *J. Petrol.* **46**, 829–858 (2005).
17. Dupré, B. & Allègre, C. J. Pb–Sr isotope variation in Indian Ocean basalts and mixing phenomena. *Nature* **303**, 142–146 (1983).
18. Hawkesworth, C. J., Mantovani, M. S. M., Taylor, P. N. & Palacz, Z. Evidence from the Paraíba of south Brazil for a continental contribution to Dupal basalts. *Nature* **322**, 356–359 (1986).
19. Castillo, P. The Dupal anomaly as a trace of the upwelling lower mantle. *Nature* **336**, 667–670 (1988).
20. Arndt, N. T. & Goldstein, S. L. An open boundary between lower continental crust and mantle: its role in crust formation and crustal recycling. *Tectonophysics* **161**, 201–212 (1989).
21. le Roex, A. P., Dick, H. J. B. & Fisher, R. L. Petrology and geochemistry of MORB from 25°E to 46°E along the Southwest Indian Ridge: Evidence for contrasting styles of mantle enrichment. *J. Petrol.* **30**, 947–986 (1989).
22. Mahoney, J. J. *et al.* Isotopic and geochemical provinces of the Western Indian Ocean spreading centers. *J. Geophys. Res.* **94**, 4033–4052 (1989).
23. Barling, J., Goldstein, S. L. & Nicholls, I. A. Geochemistry of Heard Island (southern Indian Ocean): Characterization of an enriched mantle component and implications for the enrichment of the sub-Indian Ocean mantle. *J. Petrol.* **35**, 1017–1053 (1994).
24. Rehkamper, M. & Hofmann, A. W. Recycled ocean crust and sediment in Indian Ocean MORB. *Earth Planet. Sci. Lett.* **147**, 93–106 (1997).
25. Kempton, P. D. *et al.* Sr–Nd–Pb–Hf isotope results from ODP Leg 187: Evidence for mantle dynamics of the Australian–Antarctic Discordance and origin of the Indian MORB source. *Geochem. Geophys. Geosyst.* **3**, doi:10.1029/2002GC000320 (2002).
26. Hanan, B. B., Blichert-Toft, J., Pyle, D. G. & Christie, D. M. Contrasting origins of the upper mantle revealed by hafnium and lead isotopes from the Southeast Indian Ridge. *Nature* **432**, 91–94 (2004).
27. Escriu, S., Capmas, F., Dupré, B. & Allègre, C. J. Osmium isotopic constraints on the nature of the DUPAL anomaly from Indian mid-ocean-ridge basalts. *Nature* **431**, 59–63 (2004).
28. Meyzen, C. M. *et al.* New insights into the origin and distribution of the DUPAL isotope anomaly in the Indian Ocean mantle from MORB of the Southwest Indian Ridge. *Geochem. Geophys. Geosyst.* **6**, doi:10.1029/2005GC000979 (2005).
29. Geldmacher, J., Hoernle, K., Klugel, A., van den Bogaard, P. & Bindemann, I. Geochemistry of a new enriched mantle type locality in the northern hemisphere: Implications for the origin of the EM-I source. *Earth Planet. Sci. Lett.* **265**, 167–182 (2008).
30. le Roux, P. J. *et al.* Mantle heterogeneity beneath the southern Mid-Atlantic Ridge: trace element evidence for contamination of ambient asthenospheric mantle. *Earth Planet. Sci. Lett.* **203**, 479–498 (2002).
31. Janney, P. E., le Roex, A. P. & Carlson, R. W. Hafnium isotope and trace element constraints on the nature of mantle heterogeneity beneath the central Southwest Indian Ridge (13°E to 47°E). *J. Petrol.* **46**, 2427–2464 (2005).
32. Salters, V. J. M. & Stracke, A. Composition of the depleted mantle. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000597 (2004).
33. Workman, R. K. & Hart, S. R. Major and trace element composition of the depleted MORB mantle (DMM). *Earth Planet. Sci. Lett.* **231**, 53–72 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. le Roex for comments that helped to improve this paper. This work was supported by the US National Science Foundation. G.S. was supported partly by a Paul and Daisy Soros Fellowship for New Americans.

**Author Information** The geochemical data reported here are available in the Petrological Database of the Ocean Floor ([www.petdb.org](http://www.petdb.org)). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.L.G. ([steveg@ldeo.columbia.edu](mailto:steveg@ldeo.columbia.edu)).



## METHODS

**Major elements.** Major elements were measured at the American Museum of Natural History on the AMNH-LDEO Cameca electron microprobe. Each analysis represents an average of five individual measurements on a single glass chip. A glass chip of the Lamont glass probe standard, JDF-D2, was repeatedly analysed in each probe session to monitor long-term drift. No long-term drift was evident in any run. The errors are based on the reproducibility of JDF analyses over multiple probe sessions and are  $\leq 1\%$  for  $\text{SiO}_2$  and  $\text{MgO}$ .

**Trace elements.** Hand-picked glass separates with minimal surface alteration were analysed for trace element compositions and measured on the Plasma Quad 2 inductively coupled plasma mass spectrometer at LDEO or a Thermo Electron X-series inductively coupled plasma mass spectrometer at Harvard. Solutions were spiked with Ge, In, Tm and Bi to correct for in-run drift. Concentrations were determined using a calibration curve based on US Geological Survey and in-house standards measured during each run. Errors are 3–5%.

**Isotope ratios.** Pb, Nd and Sr isotope ratios were measured on separate glass chips from the trace element analyses. These were leached cold for 12 minutes with 8N  $\text{HNO}_3$ . Pb was separated using AG1-X8 anion resin, Sr was separated using Eichrom Sr resin and Nd was separated in a two-column procedure using Eichrom TRU-spec resin to separate the rare-earth elements, followed by  $\alpha$ -hydroxy isobutyric acid. Sr, Nd and Pb isotopes were measured on a VG Sector 54 multicollector thermal ionization mass spectrometer at LDEO.  $^{87}\text{Sr}/^{86}\text{Sr}$  ratios were normalized to  $^{86}\text{Sr}/^{88}\text{Sr} = 0.1194$ . Sr isotopes were measured in multidynamic mode. Instrumental reproducibility was monitored by repeated measurements of the US National Institute of Standards and Technology SRM 987 standard, which yielded  $^{87}\text{Sr}/^{86}\text{Sr} = 0.710246 \pm 0.000016$  (2-s.d. external reproducibility,  $n > 40$ ). All samples were further corrected to a value of 0.71024 for SRM 987. All Nd analyses were made in multidynamic mode as  $\text{NdO}^+$ .  $^{143}\text{Nd}/^{144}\text{Nd}$  ratios were normalized to  $^{146}\text{Nd}/^{144}\text{Nd} = 0.7129$ . Repeated measurements of the La Jolla Nd standard yielded  $^{143}\text{Nd}/^{144}\text{Nd} = 0.511838 \pm 0.000012$  (2-s.d. external reproducibility,  $n > 30$ ). All samples were further corrected to a value of 0.511860 for the La Jolla standard. Pb isotope analyses used a  $^{207}\text{Pb}$ – $^{204}\text{Pb}$  double spike. All measurements were performed in static mode. Measurements of unspiked–double-spiked pairs of the NBS 981 standard were replicated to 183, 284 and 300 p.p.m. (2-s.d. external reproducibility,  $n = 20$ ) for  $^{206}\text{Pb}/^{204}\text{Pb}$ ,  $^{207}\text{Pb}/^{204}\text{Pb}$  and  $^{208}\text{Pb}/^{204}\text{Pb}$  ratios, respectively. These measured Pb isotope ratios were corrected to the Todt *et al.*<sup>34</sup> values of 16.9356, 15.4891, and 36.7006, respectively, for NBS 981.

34. Todt, W., Cliff, R. A., Hanser, A. & Hofmann, A. W. in *Earth Processes: Reading the Isotopic Code* Vol. 95 (eds Basu, A. & Hart, S. R.) 429–437 (American Geophysical Union, Washington DC, 1996).

## LETTERS

# Hydatellaceae are water lilies with gymnospermous tendencies

William E. Friedman<sup>1</sup>

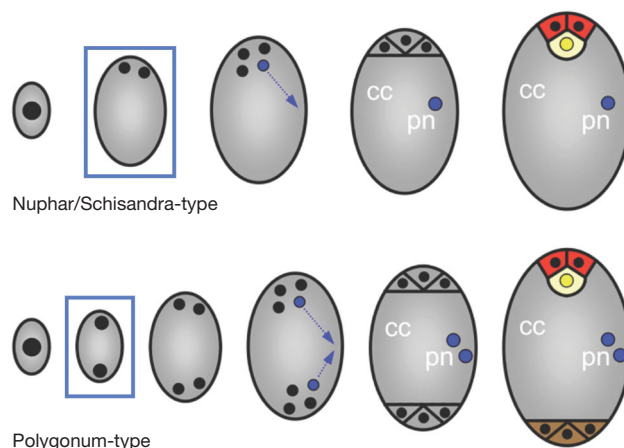
The flowering plant family Hydatellaceae was recently discovered to be allied to the ancient angiosperm lineage Nymphaeales (water lilies)<sup>1</sup>. Because of its critical phylogenetic position, members of the Hydatellaceae have the potential to provide insights into the origin and early diversification of angiosperms<sup>2</sup>. Here I report that *Hydatella* expresses several rare embryological features that, in combination, are found only in members of the Nymphaeales. At maturity, the female gametophyte is four-celled, four-nucleate and will produce a diploid endosperm, as is characteristic of most early divergent angiosperm lineages<sup>3,4</sup>. As with all members of the Nymphaeales, endosperm in *Hydatella* is minimally developed and perisperm is the major embryo-nourishing tissue within the seed<sup>5,6</sup>. Remarkably, *Hydatella* exhibits a maternal seed-provisioning strategy that is unique among flowering plants, but common to all gymnosperms<sup>7</sup>: pre-fertilization allocation of nutrients to the embryo-nourishing tissue. This exceptional case of pre-fertilization maternal provisioning of a seed in *Hydatella* may well be an apomorphic feature of Hydatellaceae alone but, given the newly discovered phylogenetic position of this family, potentially represents a plesiomorphic and transitional condition associated with the origin of flowering plants from gymnospermous ancestors.

For over a century, the flowering plant family Hydatellaceae was thought to belong to the Poales, a highly derived monocot order that includes the grasses<sup>1,2</sup>. New evidence derived from DNA sequencing and phylogenetic analysis shows that this obscure group of minute aquatic plants is closely related to water lilies (Nymphaeales)<sup>1</sup>, which themselves were only recently discovered to be one of the most ancient extant lineages of angiosperms<sup>8–12</sup>. As a result of these surprising phylogenetic insights, our understanding of the earliest phases of the radiation of angiosperms continues to be very much in flux<sup>3,13</sup>.

Although the data are, at best, ambiguous, the original embryological studies of Hydatellaceae<sup>5,14,15</sup> suggested that the female gametophyte might contain eight nuclei and seven cells, with antipodals that degenerate early (or are absent), and possibly two polar nuclei; as such, Hydatellaceae was thought to produce a triploid genetically biparental endosperm. Recently, however, members of the Nymphaeales and Austrobaileyales (another ancient lineage of flowering plants) have been shown to produce an unusual four-celled, four-nucleate female gametophyte (Nuphar/Schisandra-type) that lacks antipodal cells and a second polar nucleus (Fig. 1). Importantly, the target of the second fertilization event, the central cell of the female gametophyte, is haploid in Nymphaeales and Austrobaileyales, and endosperm in these ancient lineages is diploid, not triploid as in most flowering plants<sup>3</sup>. Because of the recent phylogenetic insights into what constitute the most ancient angiosperm lineages<sup>1</sup>, Hydatellaceae might be predicted to share some of the rare and potentially conservative embryological features of water lilies and other ancient lineages of flowering plants.

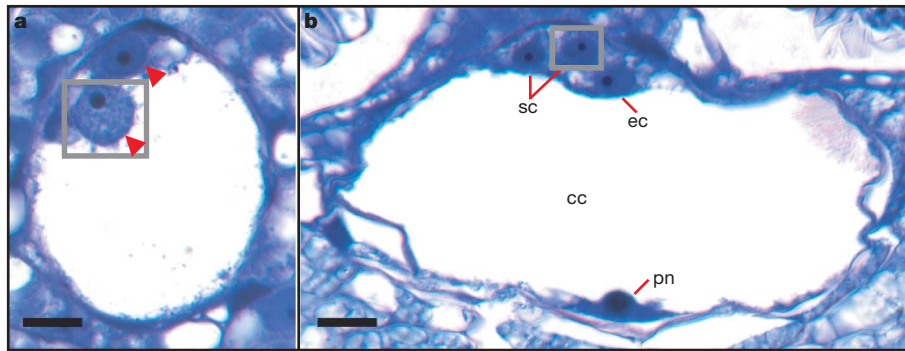
Working with field-collected material of *Hydatella inconspicua*, I examined female gametophyte development and ontogenetic features of the formation of embryo-nourishing reserves in the ovule/seed. The mature female gametophyte of *H. inconspicua* contains four uninucleate cells: the egg cell, two synergids and a uninucleate central cell (Fig. 2). No antipodal cells are evident, nor is a second polar nucleus present within the central cell. However, in many flowering plants with seven-celled, eight-nucleate female gametophytes (Polygonum-type), the three antipodal cells degenerate and the two haploid nuclei of the central cell fuse before the fertilization process<sup>3</sup>. The result is a female gametophyte that is structurally indistinguishable, at maturity, from a 'true' four-celled female gametophyte (Nuphar/Schisandra-type) that never forms antipodal cells or a second polar nucleus. To distinguish between these two types of female gametophyte, early developmental stages must be examined (Fig. 1).

Critically, at the two-nucleate syncytial stage in *H. inconspicua*, both nuclei are located at the micropylar pole of the female gametophyte (Fig. 2a). This is precisely the pattern exhibited by other four-nucleate, four-celled gametophytes found in Nymphaeales and Austrobaileyales, and differs from the comparable ontogenetic stage of Polygonum-type (and Amborella-type<sup>13</sup>) gametophytes where the nuclei at the two-nucleate syncytial stage are displaced



**Figure 1 | Schematic of four-celled, four-nucleate (Nuphar/Schisandra-type) female gametophyte development in Nymphaeales and Austrobaileyales and of seven-celled, eight-nucleate (Polygonum-type) development in most other angiosperms.** The micropylar pole is towards the top of the figure. At the two-nucleate syncytial stage (blue boxes), water-lily female gametophytes have both nuclei at the micropylar pole; in Polygonum-type female gametophytes, a nuclear migration event leads to placement of a single nucleus at each pole. Red, synergids; yellow, egg cell; brown, antipodals. cc, central cell. pn, polar nucleus.

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA.



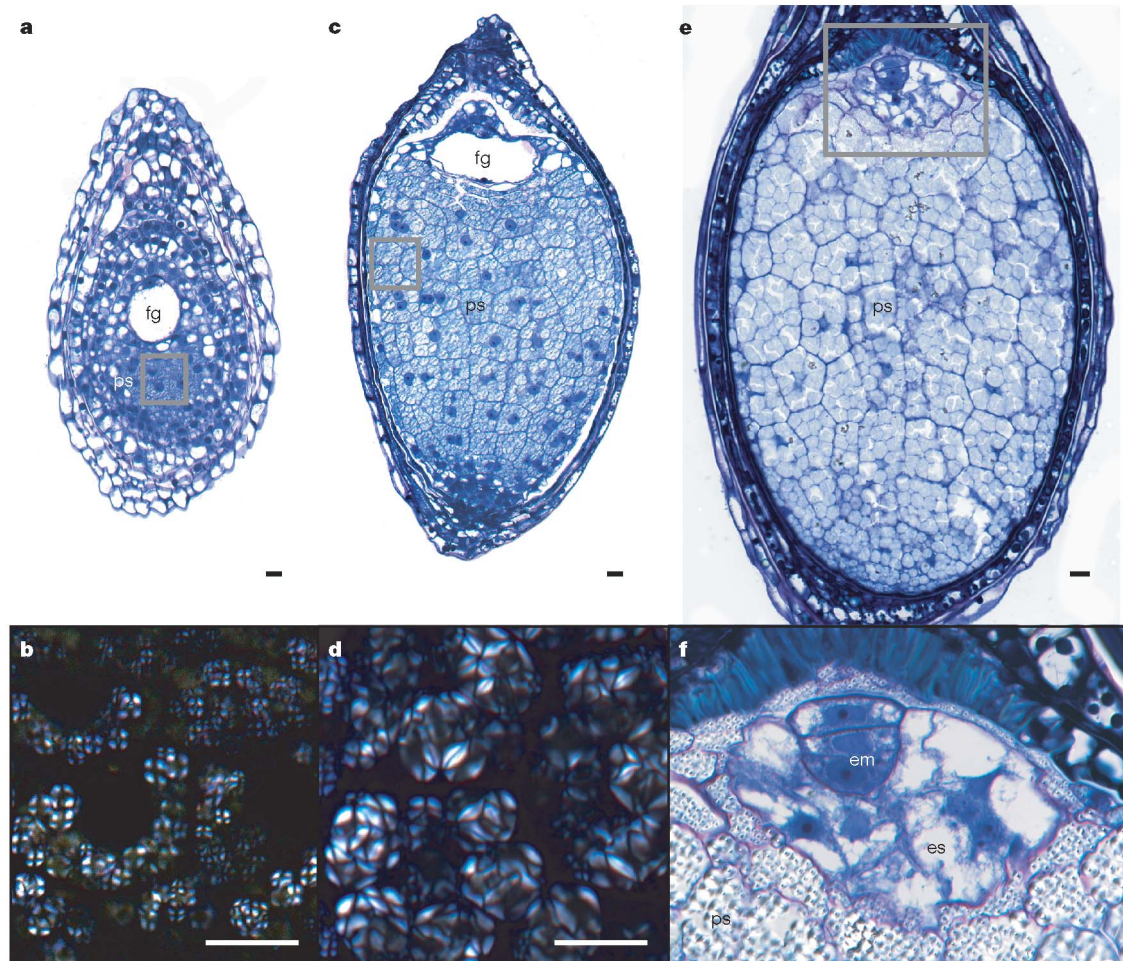
**Figure 2 | Female gametophyte development in *Hydatella inconspicua*.**

**a**, Two-nucleate syncytial female gametophyte in *H. inconspicua*, with both nuclei (red arrowheads) at micropylar pole. Grey box contains digital superposition of second nucleus from adjacent histological section.

**b**, Mature four-celled female gametophyte in *H. inconspicua*. Grey box contains digital superposition of second synergid nucleus from adjacent histological section. cc, central cell; ec, egg cell; pn, polar nucleus; sc, synergid cell.

to opposite poles of the female gametophyte (Fig. 1). Thus, *Hydatella* produces a four-celled, four-nucleate female gametophyte that precisely matches (developmentally and at maturity) the unique and potentially plesiomorphic pattern found in two of the most ancient extant angiosperm lineages (Fig. 2c)<sup>16,17</sup>. The finding that

*Hydatella* produces a female gametophyte identical to those found in the water lilies is clearly concordant with both the phylogenetic placement of Hydatellaceae as sister to the previously recognized Nymphaeales and its general position among ancient extant angiosperm clades.



**Figure 3 | Pre- and post-fertilization development of perisperm in *H. inconspicua*.** **a**, Longitudinal section of perisperm in pre-fertilization ovule at two-nucleate stage of female gametophyte development. **b**, Cross-polarization optical image demonstrating presence of starch from boxed region in **a**. Starch grains, which are birefringent, appear as four white quadrants separated by a black cross. **c**, Longitudinal section of perisperm in pre-fertilization ovule just before fertilization. **d**, Cross-polarization optical image demonstrating presence of large amounts of starch from boxed region

in **c**. Starch grains, which are birefringent, appear as four white quadrants separated by a black cross. **e**, Longitudinal section of perisperm in post-fertilization seed with three-celled embryo and minimally developed endosperm. **f**, Higher magnification from boxed region in **e** of embryo surrounded by a small endosperm tissue that lacks significant storage reserves. The perisperm below the endosperm contains large amounts of starch (bright circular structures). Scale bars, 10  $\mu$ m. em, embryo; es, endosperm; fg, female gametophyte; ps, perisperm.



Although endosperm is initiated in *H. inconspicua* (Fig. 3), perisperm, a diploid tissue derived from the maternal sporophyte, is the major embryo-nourishing constituent within the seed<sup>5,15,18</sup>, as is characteristic of all members of the Nymphaeales<sup>2,19,20</sup>. The sexually formed endosperm ultimately occupies a very small portion of the maturing seed (Fig. 3) and does not play a significant role in the nourishment of the embryo.

Remarkably, the embryo-nourishing tissue within the ovule/seed in *Hydatella* begins to acquire significant carbon resources (starch) from the maternal plant before fertilization. At the two-nucleate stage of pre-fertilization female gametophyte development, the perisperm contains reserves of starch (Fig. 3). At the four-nucleate, four-celled stage of female gametophyte development, just before fertilization, cells of the perisperm are densely packed with starch grains. After fertilization (Supplementary Data) the seed continues to develop and the volume of the perisperm increases (Fig. 3). Overall seed size increases from approximately 450 µm in length by 235 µm in width at the time of fertilization to 620 µm in length by 420 µm in width at maturity. Thus in *Hydatella* a significant portion of the maternal commitment of embryo-nourishing carbon reserves to ovules/seeds occurs before fertilization.

Angiosperms (with the exception of Hydatellaceae) differ from extant gymnosperms (conifers, *Ginkgo*, cycads, Gnetales) in that maternal commitment of embryo-nourishing resources to seeds only occurs after fertilization<sup>21,22</sup>. This ontogenetic delay in seed provisioning (compared with ancestral non-flowering seed plants) has long been viewed as an evolved and adaptive mechanism to allow the maternal plant to allocate limited resources efficiently only to those seeds that have been successfully fertilized. As such, post-fertilization maternal resource allocation has been hypothesized to be a key innovation associated with the origin and subsequent radiation of flowering plants<sup>22</sup>.

The exceptional case of pre-fertilization maternal provisioning of a seed in *Hydatella* may well be an apomorphic feature of Hydatellaceae alone. If so, Hydatellaceae stands as an essentially unique angiosperm clade that has reverted to allocating resources to an ovule/seed before the initiation of an embryo (albeit to the perisperm and not to the female gametophyte, as in gymnosperms). Because most taxa of flowering plants use a sexually formed endosperm to nourish an embryo, pre-fertilization allocation of embryo-nourishing reserves to the seed is ontogenetically precluded. However, members of a large number of disparate clades of angiosperms, including many basal lineages (Nymphaeaceae, Cabombaceae, Trimeniaceae, Acoraceae, Ceratophyllaceae, Saururaceae, Piperaceae, Hydnoraceae), form an embryo-nourishing perisperm<sup>23–28</sup>. None of these diverse perisperm-forming lineages (with the possible exception of *Acorus*) allocates significant carbon resources to the ovule/seed before fertilization<sup>23–29</sup>. Importantly, if pre-fertilization provisioning of embryo-nourishing reserves should prove to be an apomorphy of Hydatellaceae, it is more strong proof that the earliest phases of flowering plant evolution were marked by a tremendous diversification of reproductive features.

Alternatively, pre-fertilization maternal resource allocation to ovules/seeds, and specifically to a maternally derived perisperm, in *Hydatella* could represent a plesiomorphic and transitional condition associated with the origin of flowering plants. If so, the 'under-developed' endosperm of Hydatellaceae and Nymphaeales is not reduced, but rather represents an intermediate condition (between gymnosperms and other angiosperms) in which the endosperm has not yet achieved its fully fledged role as the primary source of nutrients for the developing embryo. It is important to note that *Amborella* and most members of the Austrobaileyales contain a well-developed endosperm<sup>20</sup> and do not form a perisperm (*Trimenia* being the reported exception<sup>27</sup>). Nevertheless, the prospect that early angiosperms might have used both a perisperm and an endosperm to nourish the embryo within a seed (as previously hypothesized<sup>24,30</sup>), and that the maternal plant allocated reserves to this perisperm before

fertilization, is thoroughly congruent with the data derived from *Hydatella* and its new-found phylogenetic position.

Charles Darwin was among the first to recognize the immense chasm between gymnosperms and angiosperms with respect to their biological characteristics (letters to Oswald Heer in 1875 and to Joseph Hooker in 1879 and 1882). Since then, inferring the vegetative, floral and reproductive features that defined the first angiosperms, as well as their evolutionary (transformational) links to a gymnospermous ancestor, has proved to be fraught with difficulties. Indeed, the past five years have witnessed the near-global collapse of a century-old set of paradigms concerning the embryological features of the earliest angiosperms<sup>3,13,17,20</sup>. With the present finding that maternal plants of *Hydatella* provision ovules/seeds with embryo-nourishing reserves before fertilization, yet another long-standing hypothesis for a presumed angiosperm-defining biological feature appears poised to be overturned.

## METHODS SUMMARY

*Hydatella inconspicua* was collected on 6 December 2006 at Kai Iwi Lake, Northland, New Zealand, by P. Champion, New Zealand National Institute of Water and Atmospheric Research.

Plants were chemically fixed in 4% glutaraldehyde, washed in phosphate buffer and stored in water. Specimens were dehydrated through an ethanol series, then infiltrated and embedded in glycol methacrylate (JB-4 embedding kit, Electron Microscopy Sciences). Embedded flowers were serially sectioned into 4-µm thick ribbons. Sectioned flowers were stained with 0.1% toluidine blue and examined under brightfield and cross-polarization conditions. Digital imaging was on a Zeiss Axiocam digital camera using brightfield and cross polarization optics. Images were processed with Adobe Photoshop CS2. Image manipulations were restricted to operations applied to the entire image, except as noted in specific figure legends.

Received 19 November 2007; accepted 15 January 2008.

Published online 19 March 2008.

1. Saarela, J. M. et al. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* **446**, 312–315 (2007).
2. Rudall, P. J. et al. Morphology of Hydatellaceae, an anomalous aquatic family recently recognized as an early-divergent angiosperm lineage. *Am. J. Bot.* **94**, 1073–1092 (2007).
3. Williams, J. H. & Friedman, W. E. Identification of diploid endosperm in an early angiosperm lineage. *Nature* **415**, 522–525 (2002).
4. Tobe, H., Kimoto, Y. & Prakash, N. Development and structure of the female gametophyte in *Austrobaileya scandens* (Austrobaileyaceae). *J. Plant Res.* **120**, 431–436 (2007).
5. Hamann, U. Neue Untersuchungen zur Embryologie und Systematik der Centrolepidaceae. *Bot. Jahrb.* **96**, 154–191 (1975).
6. Johri, B. M., Ambegaokar, K. B. & Srivastava, P. S. *Comparative Embryology of Angiosperms*. (McGraw-Hill, New York, 2002).
7. Friedman, W. E. & Carmichael, J. S. Heterochrony and developmental innovation: evolution of female gametophyte ontogeny in *Gnetum*, a highly apomorphic seed plant. *Evolution* **52**, 1016–1030 (1998).
8. Mathews, S. & Donoghue, M. J. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**, 947–950 (1999).
9. Parkinson, C. L., Adams, K. L. & Palmer, J. D. Multigene analyses identify the three earliest lineages of extant flowering plants. *Curr. Biol.* **9**, 1485–1488 (1999).
10. Qiu, Y.-L. et al. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
11. Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a research tool for comparative biology. *Nature* **402**, 402–404 (1999).
12. Graham, S. W. & Olmstead, R. G. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am. J. Bot.* **87**, 1712–1730 (2000).
13. Friedman, W. E. Embryological evidence for developmental lability during early angiosperm evolution. *Nature* **441**, 337–340 (2006).
14. Hamann, U. Hydatellaceae – a new family of Monocotyledoneae. *N.Z. J. Bot.* **14**, 193–196 (1976).
15. Hamann, U. in *Families and Genera of Vascular Plants* vol. IV (ed. Kubitzki, K.) 231–235 (Springer, Berlin, 1998).
16. Friedman, W. E., Gallup, W. N. & Williams, J. H. Female gametophyte development in *Kadsura*: implications for Schisandraceae, Austrobaileyales, and the early evolution of flowering plants. *Int. J. Plant Sci.* **164** (suppl.), S294–S305 (2003).
17. Friedman, W. E., Madrid, E. N. & Williams, J. H. Origin of the fittest and survival of the fittest: relating female gametophyte development to endosperm genetics. *Int. J. Plant Sci.* **169**, 79–92 (2008).

18. Rudall, P. J. The nucellus and chalaza in monocotyledons: structure and systematics. *Bot. Rev.* **63**, 140–184 (1997).
19. Schneider, E. L. Morphological studies of the Nymphaeaceae. IX. The seed of *Barclaya longifolia* Wall. *Bot. Gaz.* **139**, 223–230 (1978).
20. Floyd, S. K. & Friedman, W. E. Developmental evolution of endosperm in basal angiosperms: evidence from *Amborella* (Amborellaceae), *Nuphar* (Nymphaeaceae), and *Illicium* (Illiciaceae). *Plant Syst. Evol.* **228**, 153–169 (2001).
21. Stebbins, G. L. *Flowering Plants: Evolution above the Species Level* (Harvard Univ. Press, Cambridge, Massachusetts, 1974).
22. Tiffney, B. H. in *Paleobotany, Paleocology, and Evolution* (ed. Niklas, K. J.) 193–230 (Praeger, New York, 1981).
23. Cocucci, A. E. Estudios en el género *Prosopanche* (Hydnoraceae). III Embriologia. *Kurtziana* **9**, 19–39 (1976).
24. Rudall, P. J. & Furness, C. A. Systematics of *Acorus*: ovule and anther. *Int. J. Plant Sci.* **158**, 640–651 (1997).
25. Floyd, S. K. & Friedman, W. E. Evolution of endosperm developmental patterns among basal flowering plants. *Int. J. Plant Sci.* **161** (suppl.), S57–S81 (2000).
26. Prakash, N. & Bak, H. K. Flower and fruit development in *Piper nigrum* L. cv. Kuching. *Malays. J. Sci.* **7**, 11–19 (1982).
27. Prakash, N. in *Plant Form and Function* (eds Bhatia, B., Shukla, A. K. & Sharma, H. L.) 207–216 (Angkor, New Delhi, 1998).
28. Johnson, D. S. On the development of *Saururus cernuus* L. *Bull. Torrey Bot. Club* **27**, 365–372 (1900).
29. Shamrov, I. I. in *Embryology of Flowering Plants* (ed. Batygina, T. B.) 169–170 (Science Publishers, Enfield, New Hampshire, 2006).
30. Doyle, J. A. in *Early Evolution of Flowers* (eds Endress, P. K. & Friis, E. M.) 7–29 (Springer, Vienna, 1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** I thank: P. Champion and A. Drinnan for collecting plant materials; S. Holloway for histological work; S. Renner for translation of the embryological studies of U. Hamaan; and P. Diggle, L. Hufford, J. Williams and R. Robichaux for feedback on this manuscript. This work was supported by a National Science Foundation Research Grant.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to W.E.F. (e-mail: [ned@colorado.edu](mailto:ned@colorado.edu)).

## LETTERS

# Hierarchical structure and the prediction of missing links in networks

Aaron Clauset<sup>1,3</sup>, Cristopher Moore<sup>1,2,3</sup> & M. E. J. Newman<sup>3,4</sup>

Networks have in recent years emerged as an invaluable tool for describing and quantifying complex systems in many branches of science<sup>1–3</sup>. Recent studies suggest that networks often exhibit hierarchical organization, in which vertices divide into groups that further subdivide into groups of groups, and so forth over multiple scales. In many cases the groups are found to correspond to known functional units, such as ecological niches in food webs, modules in biochemical networks (protein interaction networks, metabolic networks or genetic regulatory networks) or communities in social networks<sup>4–7</sup>. Here we present a general technique for inferring hierarchical structure from network data and show that the existence of hierarchy can simultaneously explain and quantitatively reproduce many commonly observed topological properties of networks, such as right-skewed degree distributions, high clustering coefficients and short path lengths. We further show that knowledge of hierarchical structure can be used to predict missing connections in partly known networks with high accuracy, and for more general network structures than competing techniques<sup>8</sup>. Taken together, our results suggest that hierarchy is a central organizing principle of complex networks, capable of offering insight into many network phenomena.

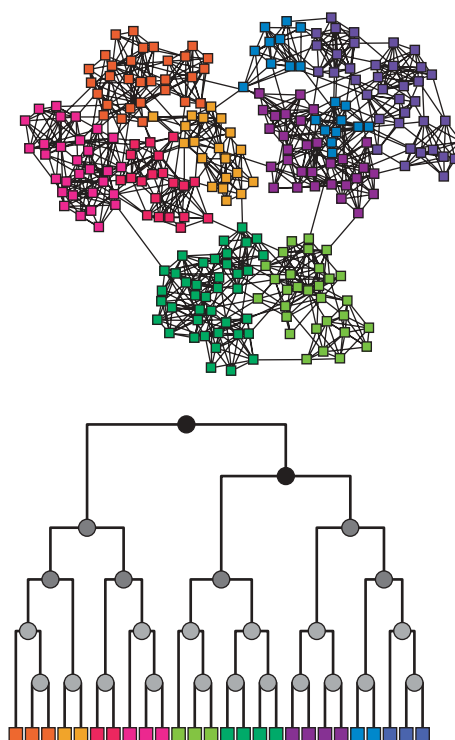
Much recent work has been devoted to the study of clustering and community structure in networks<sup>5,6,9–11</sup>. Hierarchical structure goes beyond simple clustering, however, by explicitly including organization at all scales in a network simultaneously. Conventionally, hierarchical structure is represented by a tree, or dendrogram, in which closely related pairs of vertices have lowest common ancestors that are lower in the tree than those of more distantly related pairs (see Fig. 1). We expect the probability of a connection between two vertices to depend on their degree of relatedness. Structure of this type can be modelled mathematically by using a probabilistic approach in which we endow each internal node  $r$  of the dendrogram with a probability  $p_r$  and then connect each pair of vertices for which  $r$  is the lowest common ancestor independently with probability  $p_r$  (Fig. 1).

This model, which we call a hierarchical random graph, is similar in spirit to (although different in realization from) the tree-based models used in some studies of network search and navigation<sup>12,13</sup>. Like most work on community structure, it assumes that communities at each level of organization are disjoint. Overlapping communities have occasionally been studied (see, for example, ref. 14) and could be represented with a more elaborate probabilistic model; however, as we discuss below, the present model already captures many of the structural features of interest.

Given a dendrogram and a set of probabilities  $p_r$ , the hierarchical random graph model allows us to generate artificial networks with a specified hierarchical structure, a procedure that might be useful in certain situations. Our goal here, however, is a different one. We wish

to detect and analyse the hierarchical structure, if any, of networks in the real world. We accomplish this by fitting the hierarchical model to observed network data by using the tools of statistical inference, combining a maximum-likelihood approach<sup>15</sup> with a Monte Carlo sampling algorithm<sup>16</sup> on the space of all possible dendrograms. This technique allows us to sample hierarchical random graphs with probability proportional to the likelihood that they generate the observed network. To obtain the results described below we combine information from a large number of such samples, each of which is a reasonably likely model of the data.

The success of this approach relies on the flexible nature of our hierarchical model, which allows us to fit a wide range of network structures. The traditional picture of communities or modules in a network, for example, corresponds to connections that are dense within groups of vertices and sparse between them—a behaviour called ‘assortativity’ in the literature<sup>17</sup>. The hierarchical random



**Figure 1 | A hierarchical network with structure on many scales, and the corresponding hierarchical random graph.** Each internal node  $r$  of the dendrogram is associated with a probability  $p_r$  that a pair of vertices in the left and right subtrees of that node are connected. (The shades of the internal nodes in the figure represent the probabilities.)

<sup>1</sup>Department of Computer Science, and <sup>2</sup>Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA. <sup>3</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA. <sup>4</sup>Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109, USA.



graph can capture behaviour of this kind using probabilities  $p_r$  that decrease as we move higher up the tree. Conversely, probabilities that increase as we move up the tree correspond to ‘disassortative’ structures in which vertices are less likely to be connected on small scales than on large ones. By letting the  $p_r$  values vary arbitrarily throughout the dendrogram, the hierarchical random graph can capture both assortative and disassortative structure, as well as arbitrary mixtures of the two, at all scales and in all parts of the network.

To demonstrate our method we have used it to construct hierarchical decompositions of three example networks drawn from disparate fields: the metabolic network of the spirochaete *Treponema pallidum*<sup>18</sup>, a network of associations between terrorists<sup>19</sup>, and a food web of grassland species<sup>20</sup>. To test whether these decompositions accurately capture the important structural features of the networks, we use the sampled dendrograms to generate new networks, different in detail from the originals but, by definition, having similar hierarchical structure (see Supplementary Information for more details). We find that these ‘resampled’ networks match the statistical properties of the originals closely, including their degree distributions, clustering coefficients, and distributions of shortest path lengths between pairs of vertices, despite the fact that none of these properties is explicitly represented in the hierarchical random graph (Table 1, and Supplementary Fig. 3). It therefore seems that a network’s hierarchical structure is capable of explaining a wide variety of other network features as well.

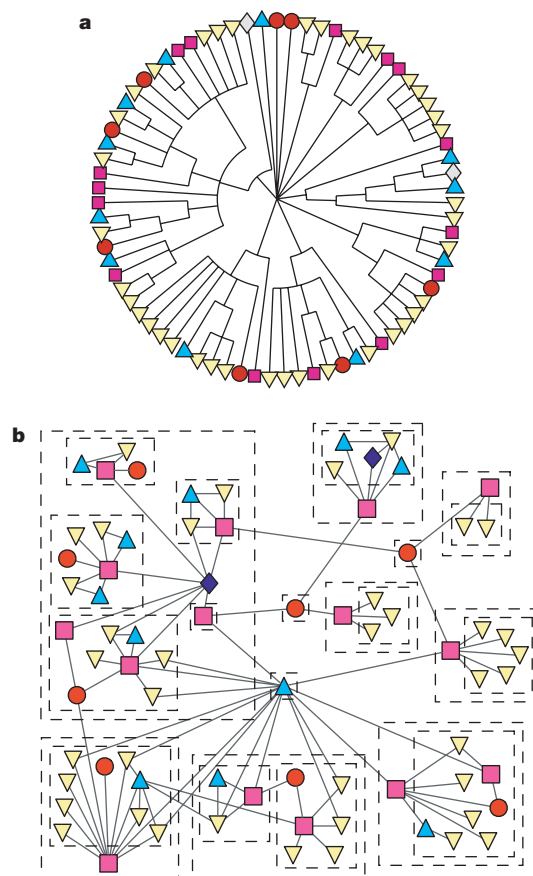
The dendrograms produced by our method are also of interest in themselves, as a graphical representation and summary of the hierarchical structure of the observed network. As discussed above, our method can generate not just a single dendrogram but a set of dendrograms, each of which is a good fit to the data. From this set we can, by using techniques from phylogeny reconstruction<sup>21</sup>, create a single consensus dendrogram, which captures the topological features that appear consistently across all or a large fraction of the dendrograms and typically is a better summary of the network’s structure than any individual dendrogram. Figure 2a shows such a consensus dendrogram for the grassland species network, which clearly reveals communities and subcommunities of plants, herbivores, parasitoids and hyperparasitoids.

Another application of the hierarchical decomposition is the prediction of missing interactions in networks. In many settings, the discovery of interactions in a network requires significant experimental effort in the laboratory or the field. As a result, our current pictures of many networks are substantially incomplete<sup>22–28</sup>. An alternative to checking exhaustively for a connection between every pair of vertices in a network is to try to predict, in advance and on the basis of the connections already observed, which vertices are most likely to be connected, so that scarce experimental resources can be focused on testing for those interactions. If our predictions are good, we can in this way substantially reduce the effort required to establish the network’s topology.

The hierarchical decomposition can be used as the basis for an effective method of predicting missing interactions as follows. Given an observed but incomplete network, we generate, as described above, a set of hierarchical random graphs—dendrograms and the associated probabilities  $p_r$ —that fit that network. Then we look for pairs of vertices that have a high average probability of connection within these hierarchical random graphs but are unconnected in the

observed network. These pairs we consider the most likely candidates for missing connections. (Technical details of the procedure are given in Supplementary Information.)

We demonstrate the method by using our three example networks again. For each network we remove a subset of connections chosen uniformly at random and then attempt to predict, on the basis of the remaining connections, which have been removed. A standard metric for quantifying the accuracy of prediction algorithms, commonly used in the medical and machine learning communities, is the AUC statistic, which is equivalent to the area under the receiver operating characteristic (ROC) curve<sup>29</sup>. In the present context, the AUC statistic can be interpreted as the probability that a randomly chosen missing connection (a true positive) is given a higher score by our method than a randomly chosen pair of unconnected vertices (a true negative). Thus, the degree to which the AUC exceeds 0.5 indicates how much better our predictions are than chance. Figure 2 shows the AUC statistic for the three networks as a function of the fraction of the connections known to the algorithm. For all three networks our algorithm does far better than chance, indicating that hierarchy is a strong general predictor of missing structure. It is also instructive to compare the performance of our method with that of other methods for link prediction<sup>8</sup>. Previously proposed methods include assuming that vertices are likely to be connected if they have many common neighbours, if there are short paths between them, or if the product of their degrees is large. These approaches work well for strongly assortative networks such as collaboration and citation

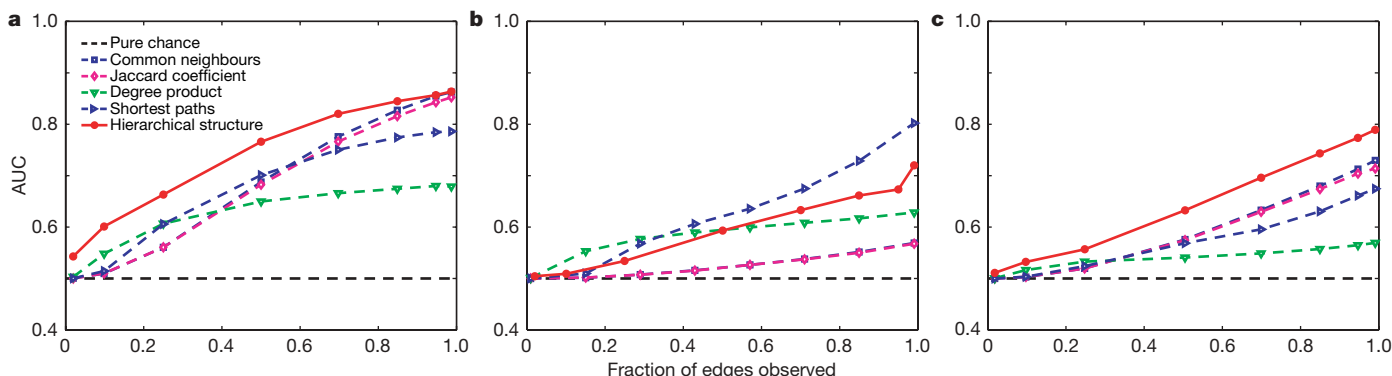


**Figure 2 | Application of the hierarchical decomposition to the network of grassland species interactions.** **a**, Consensus dendrogram reconstructed from the sampled hierarchical models. **b**, A visualization of the network in which the upper few levels of the consensus dendrogram are shown as boxes around species (plants, herbivores, parasitoids, hyperparasitoids and hyperparasitoids are shown as circles, boxes, down triangles, up triangles and diamonds, respectively). Note that in several cases a set of parasitoids is grouped into a disassortative community by the algorithm, not because they prey on each other but because they prey on the same herbivore.

**Table 1 | Comparison of original and resampled networks**

Network	$\langle k \rangle_{\text{real}}$	$\langle k \rangle_{\text{samp}}$	$C_{\text{real}}$	$C_{\text{samp}}$	$d_{\text{real}}$	$d_{\text{samp}}$
<i>T. pallidum</i>	4.8	3.7(1)	0.0625	0.0444(2)	3.690	3.940(6)
Terrorists	4.9	5.1(2)	0.361	0.352(1)	2.575	2.794(7)
Grassland	3.0	2.9(1)	0.174	0.168(1)	3.29	3.69(2)

Statistics are shown for the three example networks studied and for new networks generated by resampling from our hierarchical model. The generated networks closely match the average degree ( $\langle k \rangle$ ), clustering coefficient  $C$  and average vertex–vertex distance  $d$  in each case, suggesting that they capture much of the structure of the real networks. Parenthetical values indicate standard errors on the final digits.



**Figure 3 | Comparison of link prediction methods.** Average AUC statistic—that is, the probability of ranking a true positive over a true negative—as a function of the fraction of connections known to the algorithm, for the link

prediction method presented here and a variety of previously published methods. **a**, Terrorist association network; **b**, *T. pallidum* metabolic network; **c**, grassland species network.

networks<sup>8</sup> and for the metabolic and terrorist networks studied here (Fig. 3a, b). Indeed, for the metabolic network the shortest-path heuristic performs better than our algorithm.

However, these simple methods can be misleading for networks that exhibit more general types of structure. In food webs, for instance, pairs of predators often share prey species but rarely prey on each other. In such situations a common-neighbour or shortest-path-based method would predict connections between predators where none exists. The hierarchical model, by contrast, is capable of expressing both assortative and disassortative structure and, as Fig. 3c shows, gives substantially better predictions for the grassland network. (Indeed, in Fig. 2b there are several groups of parasitoids that our algorithm has grouped together in a disassortative community, in which they prey on the same herbivore but not on each other.) The hierarchical method thus makes accurate predictions for a wider range of network structures than the previous methods.

In the applications above, we have assumed for simplicity that there are no false positives in our network data; that is, that every observed edge corresponds to a real interaction. In networks in which false positives may be present, however, they too could be predicted by using the same approach: we would simply look for pairs of vertices that have a low average probability of connection within the hierarchical random graph but are connected in the observed network.

The method described here could also be extended to incorporate domain-specific information, such as species' morphological or behavioural traits for food webs<sup>28</sup> or phylogenetic or binding-domain data for biochemical networks<sup>23</sup>, by adjusting the probabilities of edges accordingly. As the results above show, however, we can obtain good predictions even in the absence of such information, indicating that topology alone can provide rich insights.

In closing, we note that our approach differs crucially from previous work on hierarchical structure in networks<sup>1,4–7,9,11,30</sup> in that it acknowledges explicitly that most real-world networks have many plausible hierarchical representations of roughly equal likelihood. Previous work, by contrast, has typically sought a single hierarchical representation for a given network. By sampling an ensemble of dendrograms, our approach avoids over-fitting the data and allows us to explain many common topological features, to generate resampled networks with similar structure to the original, to derive a clear and concise summary of a network's structure by means of its consensus dendrogram, and to accurately predict missing connections in a wide variety of situations.

## METHODS SUMMARY

Computer code implementing many of the analysis methods described in this paper can be found online at <http://www.santafe.edu/~aaronc/randomgraphs/>.

Received 13 August 2007; accepted 7 February 2008.

1. Wasserman, S. & Faust, K. *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
2. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
3. Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
4. Ravasz, E., Somera, A. L., Mongru, D. A., Olvtai, Z. N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **30**, 1551–1555 (2002).
5. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004).
6. Guimera, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
7. Lagomarsino, M. C., Jona, P., Bassetti, B. & Isambert, H. Hierarchy and feedback in the evolution of the *Escherichia coli* transcription network. *Proc. Natl Acad. Sci. USA* **104**, 5516–5520 (2001).
8. Liben-Nowell, D. & Kleinberg, J. M. The link-prediction problem for social networks. *J. Am. Soc. Inform. Sci. Technol.* **58**, 1019–1031 (2007).
9. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
10. Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food-web structure. *Nature* **426**, 282–285 (2003).
11. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. & Parisi, D. Defining and identifying communities in networks. *Proc. Natl Acad. Sci. USA* **101**, 2658–2663 (2004).
12. Watts, D. J., Dodds, P. S. & Newman, M. E. J. Identity and search in social networks. *Science* **296**, 1302–1305 (2002).
13. Kleinberg, J. in *Proc. 2001 Neural Inform. Processing Systems Conf.* (eds Dietterich, T. G., Becker, S. & Ghahramani, Z.) 431–438 (MIT Press, Cambridge, MA, 2002).
14. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).
15. Casella, G. & Berger, R. L. *Statistical Inference* (Duxbury, Belmont, 2001).
16. Newman, M. E. J. & Barkema, G. T. *Monte Carlo Methods in Statistical Physics* (Clarendon, Oxford, 1999).
17. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
18. Huss, M. & Holme, P. Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks. *IET Syst. Biol.* **1**, 280–285 (2007).
19. Krebs, V. Mapping networks of terrorist cells. *Connections* **24**, 43–52 (2002).
20. Dawah, H. A., Hawkins, B. A. & Claridge, M. F. Structure of the parasitoid communities of grass-feeding chalcid wasps. *J. Anim. Ecol.* **64**, 708–720 (1995).
21. Bryant, D. in *BioConsensus* (eds Janowitz, M., Lapointe, F.-J., McMorris, F. R., Mirkin, B. & Roberts, F.) pp. 163–184 (Series in Discrete Mathematics and Theoretical Computer Science, Vol. 61, American Mathematical Society-DIMACS, Providence, RI, 2003).
22. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: The role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12917–12922 (2002).
23. Szilágyi, A., Grimm, V., Arakaki, A. K. & Skolnick, J. Prediction of physical protein–protein interactions. *Phys. Biol.* **2**, S1–S16 (2005).

24. Sprinzak, E., Sattath, S. & Margalit, H. How reliable are experimental protein-protein interaction data? *J. Mol. Biol.* **327**, 919–923 (2003).
25. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
26. Lakhina, A., Byers, J. W., Crovella, M. & Xie, P. in *INFOCOM 2003: Twenty-Second Annual Joint Conf. IEEE Computer and Communications Societies* (ed. Bauer, F.) Vol. 1 332–341 (IEEE, Piscataway, New Jersey, 2003).
27. Clauset, A. & Moore, C. Accuracy and scaling phenomena in Internet mapping. *Phys. Rev. Lett.* **94**, 018701 (2005).
28. Martinez, N. D., Hawkins, B. A., Dawah, H. A. & Feifarek, B. P. Effects of sampling effort on characterization of food-web structure. *Ecology* **80**, 1044–1055 (1999).
29. Hanely, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
30. Sales-Pardo, M., Guimerà, R., Moreira, A. A. & Amaral, L. A. N. Extracting the hierarchical organization of complex systems. *Proc. Natl Acad. Sci. USA* **104**, 15224–15229 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Dunne, M. Gastner, P. Holme, M. Huss, M. Porter, C. Shalizi and C. Wiggins for their help, and the Santa Fe Institute for its support. C.M. thanks the Center for the Study of Complex Systems at the University of Michigan for hospitality while some of this work was conducted.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.C. ([aaronc@santafe.edu](mailto:aaronc@santafe.edu)).



## LETTERS

# Melanopsin cells are the principal conduits for rod–cone input to non-image-forming vision

Ali D. Güler<sup>1\*</sup>, Jennifer L. Ecker<sup>1\*</sup>, Gurprit S. Lall<sup>2\*</sup>, Shafiqul Haq<sup>3</sup>, Cara M. Altimus<sup>1</sup>, Hsi-Wen Liao<sup>3</sup>, Alun R. Barnard<sup>2</sup>, Hugh Cahill<sup>3</sup>, Tudor C. Badea<sup>4</sup>, Haiqing Zhao<sup>1</sup>, Mark W. Hankins<sup>5</sup>, David M. Berson<sup>6</sup>, Robert J. Lucas<sup>2</sup>, King-Wai Yau<sup>3</sup> & Samer Hattar<sup>1</sup>

Rod and cone photoreceptors detect light and relay this information through a multisynaptic pathway to the brain by means of retinal ganglion cells (RGCs)<sup>1</sup>. These retinal outputs support not only pattern vision but also non-image-forming (NIF) functions, which include circadian photoentrainment and pupillary light reflex (PLR). In mammals, NIF functions are mediated by rods, cones and the melanopsin-containing intrinsically photosensitive retinal ganglion cells (ipRGCs)<sup>2,3</sup>. Rod–cone photoreceptors and ipRGCs are complementary in signalling light intensity for NIF functions<sup>4–12</sup>. The ipRGCs, in addition to being directly photosensitive, also receive synaptic input from rod–cone networks<sup>13,14</sup>. To determine how the ipRGCs relay rod–cone light information for both image-forming and non-image-forming functions, we genetically ablated ipRGCs in mice. Here we show that animals lacking ipRGCs retain pattern vision but have deficits in both PLR and circadian photoentrainment that are more extensive than those observed in melanopsin knockouts<sup>8,10,11</sup>. The defects in PLR and photoentrainment resemble those observed in animals that lack phototransduction in all three photoreceptor classes<sup>6</sup>. These results indicate that light signals for irradiance detection are dissociated from pattern vision at the retinal ganglion cell level, and animals that cannot detect light for NIF functions are still capable of image formation.

The retinal ganglion cells that express melanopsin (rendering them intrinsically photosensitive) send monosynaptic projections to the suprachiasmatic nucleus (SCN) and the intergeniculate leaflet (IGL), responsible for circadian photoentrainment, and the olivary pretectal nucleus (OPN), responsible for PLR<sup>15</sup>. It has been reported that RGCs that do not contain melanopsin innervate the same NIF brain centres<sup>15–19</sup>. In mice genetically engineered to lack melanopsin protein, the RGCs that would normally express this opsin still project to the NIF centres, but these cells are no longer intrinsically photosensitive<sup>8</sup>. In these animals, PLR and circadian light responses are reduced but not absent, indicating that rods and cones are capable of light detection for NIF functions<sup>8–11</sup>. The rod–cone signals for NIF functions are undoubtedly relayed to the brain by RGCs, but it is unclear whether ipRGCs, conventional RGCs, or both are responsible (Fig. 1a).

To eliminate ipRGCs, we introduced a gene (*aDTA*) encoding attenuated diphtheria toxin A subunit (*aDTA*)<sup>20</sup> into the mouse gene locus encoding melanopsin (Supplementary Fig. 1a, b). Using antibody staining in retinas from animals expressing *aDTA* (*Opn4<sup>aDTA/+</sup>*), we found that  $3.1 \pm 1.5\%$  (mean  $\pm$  s.e.m.) of melanopsin cells remained in the *Opn4<sup>aDTA/+</sup>* animals (Fig. 1b). Staining

with 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-gal) in animals with the *aDTA* gene in one melanopsin allele and *tau-LacZ* in the other (*Opn4<sup>aDTA/tau-LacZ</sup>*)<sup>3</sup> revealed that  $17.2 \pm 1.0\%$  of melanopsin cells remained in the retinas of these animals at six months of age (Fig. 1c). Consistent with the elimination of ipRGCs, we observed fewer fibre terminals in the SCN, IGL and OPN of *Opn4<sup>aDTA/tau-LacZ</sup>* animals than in those of *Opn4<sup>tau-LacZ/+</sup>* animals (Fig. 2a–c). The degree of ipRGC ablation in the heterozygous animals increased with age (Supplementary Fig. 2a–c).

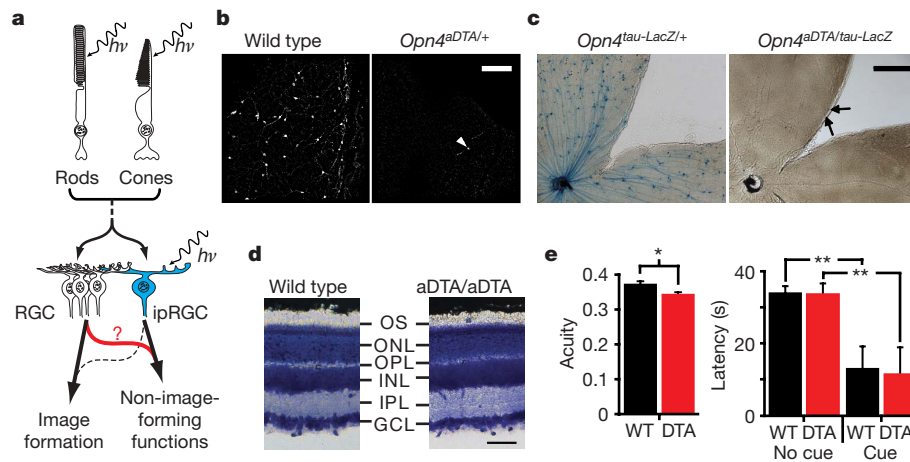
To ablate ipRGCs more completely, we generated animals homozygous for *aDTA* (*Opn4<sup>aDTA/aDTA</sup>*). The morphology and thickness of the retina in the *Opn4<sup>aDTA/aDTA</sup>* animals were not different from those in wild types (Fig. 1d). Injection of fluorescently conjugated cholera toxin into the eye, which labels all ganglion cell fibres from the retina, showed that few fibres innervated the SCN and IGL in the *Opn4<sup>aDTA/aDTA</sup>* animals (Fig. 2d, e). Furthermore, comparison between age-matched *Opn4<sup>aDTA/+</sup>* and *Opn4<sup>aDTA/aDTA</sup>* mice verified that the extent of ipRGC ablation was greater in homozygous animals (Supplementary Fig. 2d). We also observed that target innervation by other RGCs was unaffected both in the dorsal lateral geniculate nucleus, which is important for image formation, and in the rostral core of the OPN<sup>15</sup> (Fig. 2e, f).

To assess whether image-forming functions are affected in animals expressing *aDTA*, we measured electroretinograms, optokinetic nystagmus responses, visual acuity and the ability of the animals to detect a visual cue. We found that the electroretinograms and optokinetic nystagmus responses were normal in animals lacking ipRGCs (Supplementary Fig. 3a, b). On the basis of optomotor responses, the acuity of *Opn4<sup>aDTA/aDTA</sup>* mice was slightly decreased compared with that of the wild-type mice (Fig. 1e) and *Opn4<sup>tau-LacZ/tau-LacZ</sup>* mice (Supplementary Fig. 3d). This effect is most probably due to enlarged pupil diameters in the *Opn4<sup>aDTA/aDTA</sup>* animals (see Fig. 3c and Supplementary Fig. 3c). We also determined that these animals could use a visual cue to locate a flag-marked platform in a water maze (Fig. 1e). These results demonstrate that image formation is functional despite the elimination of ipRGCs in *Opn4<sup>aDTA/aDTA</sup>* animals. We could therefore determine the relative contribution of ipRGCs to PLR and circadian photoentrainment in the context of normal image formation.

Pupil constriction regulates the amount of light entering the eye, and thus pupil diameter is negatively correlated with light intensity. At high light intensity, the iris decreases the area of the pupil by 95% (full constriction) in comparison with dark-adapted conditions (fully dilated). At low light intensity in which the pupil constricts

<sup>1</sup>Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>2</sup>Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. <sup>3</sup>Department of Neuroscience, and <sup>4</sup>Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. <sup>5</sup>Visual Neuroscience, University of Oxford, Oxford OX3 7BN, UK. <sup>6</sup>Department of Neuroscience, Brown University, Providence, Rhode Island 02912, USA.

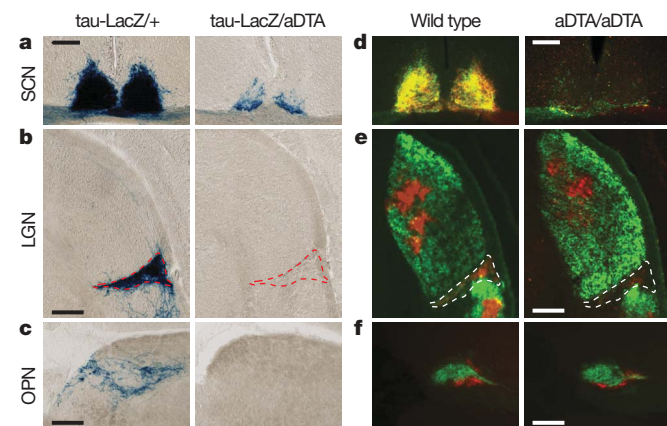
\*These authors contributed equally to this work.



**Figure 1 | Elimination of ipRGCs in mouse retina.** **a**, Model describing how rod-cone signalling through conventional RGCs or ipRGCs contributes to NIF functions. The role of ipRGCs in image formation is speculative (dotted line). **b**, Melanopsin antibody staining in retinas of 18-month-old wild-type ( $n = 6$ ) and  $Opn4^{aDTA/+}$  ( $n = 12$ ) mice. The white arrowhead indicates a surviving ipRGC. Scale bar, 200  $\mu\text{m}$ . **c**, X-gal staining from  $Opn4^{\text{tau-LacZ}/+}$  ( $n = 6$ ) and  $Opn4^{aDTA/\text{tau-LacZ}}$  ( $n = 8$ ). The surviving cells are weakly stained (black arrows). Scale bar, 500  $\mu\text{m}$ . **d**, Cross-sections of Giemsa-stained retinas from 18-month-old  $Opn4^{aDTA/aDTA}$  ( $n = 3$ ) and wild-type ( $n = 3$ ) mice. The morphology of retinas is indistinguishable between

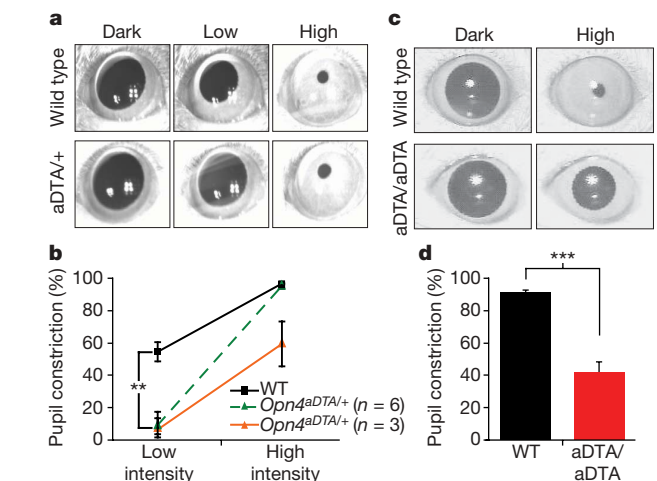
by 50% or less, rod-cone input is the main signal<sup>8,12</sup> (Supplementary Table 1). In contrast, the intrinsic photosensitivity through the melanopsin protein in ipRGCs is necessary for full pupil constriction<sup>8,12</sup> (Supplementary Table 1). We found that the PLR was absent in all  $Opn4^{aDTA/+}$  mice at a light intensity that constricts the pupil in wild-type animals to about 50% (Fig. 3a, b). This result suggests that although the melanopsin protein is not required for PLR at low light intensity, rod-cone input still requires the ipRGCs. We propose that ipRGCs are able to encode light intensities that are below the detection level of the melanopsin photopigment by integrating rod-cone signals and relaying this information to the OPN for pupil constriction.

At high light intensity, six of nine  $Opn4^{aDTA/+}$  mice constricted their pupils to 95%, in a similar manner to wild-type animals (Fig. 3a, b). Full pupil constriction in these six animals indicates that the intrinsic photosensitivity persists in the remaining ipRGCs in  $Opn4^{aDTA/+}$  mice and that only about 17% of ipRGCs are sufficient to drive full pupil constriction. The remaining three animals had



**Figure 2 | The ipRGC fibres in the brain decrease in aDTA mice.** **a–c**, X-gal staining in  $Opn4^{\text{tau-LacZ}/+}$  mice ( $Opn4^{\text{tau-LacZ}/aDTA}$ ;  $n = 2$ ) shows that ipRGC innervation of the SCN, IGL and OPN is decreased in comparison with  $Opn4^{\text{tau-LacZ}/+}$  mice ( $Opn4^{\text{tau-LacZ}/+}$ ;  $n = 2$ ). **d–f**, Ocular cholera toxin injections (left eye, green; right, red) of  $Opn4^{aDTA/aDTA}$  mice (aDTA/aDTA;  $n = 11$ ) and wild-type mice ( $n = 6$ ). **a**, **d**, SCN innervation is sparse in aDTA mice. **b**, **e**, The dorsal lateral geniculate nucleus (LGN) is innervated similarly both in aDTA and wild-type animals, whereas few fibres remain in the IGL of mutant mice (outlined regions). **c**, **f**, The OPN shell is innervated by ipRGCs and the core is targeted by other RGCs<sup>15</sup>. **c**, Fibres in the shell region are eliminated in  $Opn4^{\text{tau-LacZ}/aDTA}$  animals. **f**, Fibres in the OPN core are retained in  $Opn4^{aDTA/aDTA}$  mice. Scale bars, 200  $\mu\text{m}$ .

genotypes. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; OS, outer segment. Scale bar, 50  $\mu\text{m}$ . **e**, Left, the acuity (in cycles per degree) of  $Opn4^{aDTA/aDTA}$  mice (DTA;  $n = 11$ ; red bar) was slightly decreased in comparison with wild-type mice (WT;  $n = 9$ ; black bar). Right, the latency to locate a marked platform (cue) in a water maze was similar in  $Opn4^{aDTA/aDTA}$  (DTA;  $n = 14$ ; red bar) and wild-type (WT;  $n = 12$ ; black bar) mice. This latency significantly differed from that for unmarked platform (no cue) tests. All statistical comparisons used Student's  $t$ -test (asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ); error bars indicate s.e.m.



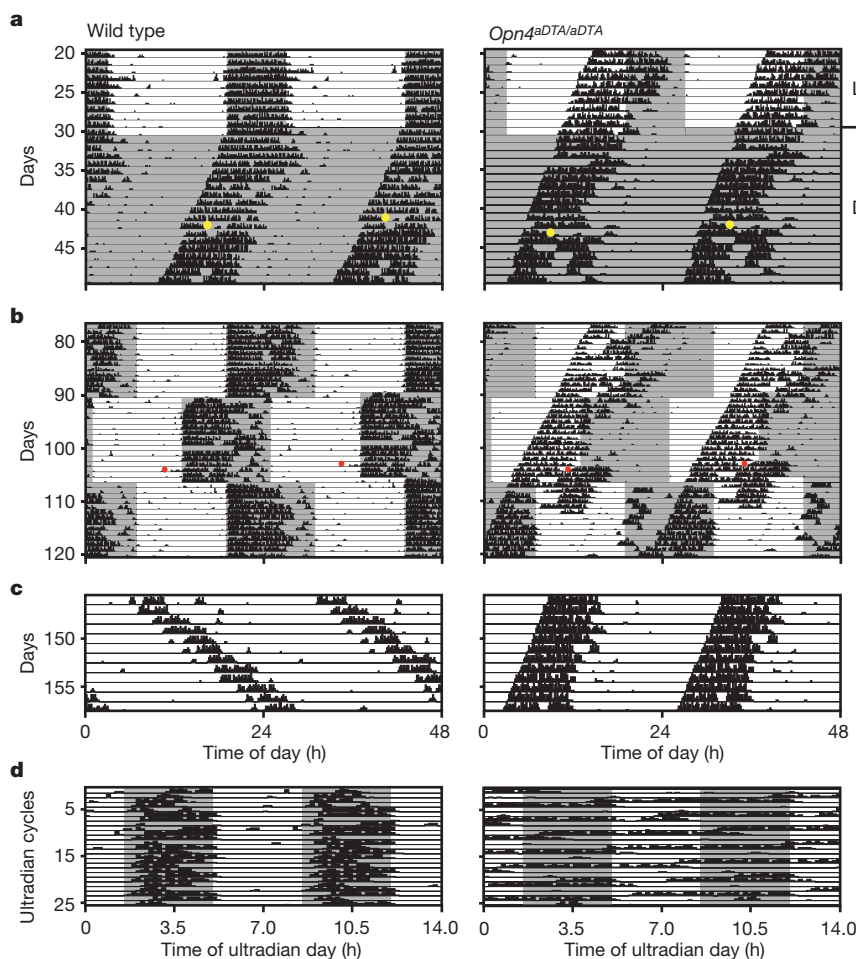
**Figure 3 |  $Opn4^{aDTA}$  mice have deficits in PLR.** **a**, All nine  $Opn4^{aDTA/+}$  (aDTA/aDTA) mice showed defective PLR at a light intensity (1.8  $\mu\text{W cm}^{-2}$ ; 30 s white light; Low) that induced 50% constriction in wild types. The  $Opn4^{aDTA/+}$  mice showed a  $9.0 \pm 6.0\%$  constriction. Six of nine  $Opn4^{aDTA/+}$  mice had full pupil constriction at high light intensity (3 mW  $\text{cm}^{-2}$ ; High). The rest of the mutant mice (three of nine) showed defective PLR. **b**, Quantification of PLR data of wild-type (WT;  $n = 9$ ; black squares) and  $Opn4^{aDTA/+}$  either photoentrained (green triangles;  $n = 6$ ) or non-photoentrained (orange triangles;  $n = 3$ ) animals. All statistical comparisons were made by Student's  $t$ -test (two asterisks,  $P < 0.01$ ). **c**, All  $Opn4^{aDTA/aDTA}$  animals (aDTA/aDTA) constrict their pupils only to a maximum of 42% after a light pulse that causes 95% constriction in wild types (161  $\mu\text{W cm}^{-2}$ , 470 nm monochromatic light; High). **d**, Quantification of PLR data of wild-type (WT;  $n = 11$ ; black bar) and  $Opn4^{aDTA/aDTA}$  ( $n = 12$ ; red bar) mice from **c**. Statistical comparisons were made by Student's  $t$ -test (three asterisks,  $P < 0.001$ ); error bars indicate s.e.m.

pupil constriction defects at high light intensity ( $59 \pm 14\%$ , in contrast with  $96 \pm 1\%$  in the wild type; Fig. 3b). Interestingly, when the *Opn4<sup>aDTA/aDTA</sup>* animals were placed under 24-h light/dark cycles, animals that had full pupil constriction photoentrained (Supplementary Fig. 4a), whereas animals with defective pupil constriction had photoentrainment defects (Supplementary Fig. 4b).

The *aDTA* homozygotes have greater ablation of ipRGCs, and consequently all 12 *Opn4<sup>aDTA/aDTA</sup>* animals had defective PLR ( $42 \pm 6\%$ ) at high light intensity (Fig. 3c, d). A similar pupil constriction (about 40%) to that observed in *Opn4<sup>aDTA/aDTA</sup>* mice is achieved by both wild-type and melanopsin knockout animals in response to a light stimulation that is about 1,000-fold lower in intensity<sup>8</sup>. This reduction in pupil sensitivity was observed in an intensity–response curve in *Opn4<sup>aDTA/aDTA</sup>* mice at constrictions that are 55% or higher in the wild types (Supplementary Fig. 5). Together, these data show that the rod–cone-dependent pupil constriction is reliant on signalling through ipRGCs at all light intensities (Fig. 3 and Supplementary Fig. 5). The residual pupil constriction observed in the homozygous animals may reveal a possible role for the non-melanopsin RGCs in pupil constriction. Alternatively, this residual pupil constriction could originate from rod–cone input through a few remaining ipRGCs in the *Opn4<sup>aDTA/aDTA</sup>* mice.

To assess the contribution of ipRGCs to circadian photoentrainment, we analysed the wheel running activity of wild-type ( $n = 11$ ) and *Opn4<sup>aDTA/aDTA</sup>* ( $n = 12$ ) animals using 24-h light/dark cycles. Under constant dark conditions (DD), both wild-type and

*Opn4<sup>aDTA/aDTA</sup>* animals had a functional circadian oscillator with period lengths of  $23.3 \pm 0.1$  and  $23.8 \pm 0.1$  h, respectively (Fig. 4a). In a similar manner to *math5<sup>-/-</sup>* mice with severe optic nerve hypoplasia due to RGC loss, *Opn4<sup>aDTA/aDTA</sup>* animals had significantly longer periods than wild types (Supplementary Table 2), confirming the loss of the fibres projecting to the SCN<sup>21</sup>. A light stimulus (1,500 lx) administered for 15 min at circadian time (CT)16 delayed the phase onset of the activity of the wild types by  $1.66 \pm 0.23$  h, whereas it did not affect the phase onset of activity in any *Opn4<sup>aDTA/aDTA</sup>* animals ( $-0.06 \pm 0.09$  h) (Fig. 4a). When the light/dark cycle was advanced or delayed by 6 h, the activity of all wild-type animals synchronized with the shifted cycle (Fig. 4b and Supplementary Table 2). *Opn4<sup>aDTA/aDTA</sup>* animals segregated into two groups with distinct responses. The first group of animals (8 of 12) free-ran completely under the shifted light/dark cycle (Fig. 4b and Supplementary Fig. 6b–f) in a similar manner to animals lacking all functional photoreceptors in the retina<sup>6</sup>, indicating that they were completely blind to the light shift. Animals in the second group (4 of 12) were weakly light responsive but not photoentrained, because they did not show a stable phase relation to the light/dark cycle (Supplementary Fig. 7). Given that these four animals demonstrated very weak light responsiveness in the light/dark cycle but were not phase-shifted by the 15-min light pulse, we tested whether constant light (LL) conditions would lengthen the circadian period of *Opn4<sup>aDTA/aDTA</sup>* mice. Under LL, 7 of 11 wild-type animals had periods longer than 24 h, whereas the remaining animals became



**Figure 4 | *Opn4<sup>aDTA/aDTA</sup>* mice do not photoentrain or mask.**

**a**, *Opn4<sup>aDTA/aDTA</sup>* mice free-run under light/dark cycles (grey and white backgrounds; dark and light (about 700 lx), respectively). *Opn4<sup>aDTA/aDTA</sup>* mice do not phase shift in response to a 15-min 1,500-lx pulse of white light (CT16; yellow dots indicate light pulses). **b**, *Opn4<sup>aDTA/aDTA</sup>* mice do not

photoentrain to the 24-h light/dark cycle in the delay or advance phases (red dots indicate cage changes). **c**, Unlike wild-type animals, no *Opn4<sup>aDTA/aDTA</sup>* mice lengthened their period under constant light. **d**, *Opn4<sup>aDTA/aDTA</sup>* mice do not mask under a 7-h ultradian cycle.



completely arrhythmic (Supplementary Fig. 6a). In contrast, both groups of *Opn4<sup>aDTA/aDTA</sup>* animals had periods shorter than 24 h (Fig. 4c, Supplementary Figs 6b–f and 7, and Supplementary Table 2). It had previously been shown that animals lacking only the melanopsin protein are similar to wild-type controls in adjusting to a 24-h light/dark cycle and have minor defects in their period lengthening under constant light and phase-delaying pulses of light<sup>10,11</sup>. The severe defects in circadian photoentrainment observed in animals lacking ipRGCs demonstrate that these cells are required not only for intrinsically signalling light information through the melanopsin protein but also for conveying rod–cone light information to the SCN. Although several reports indicated that RGCs that do not express melanopsin innervate the SCN<sup>17–19,22</sup>, our results reveal that the contribution of rod–cone signalling through these RGCs for photoentrainment is negligible.

The acute effects of light on activity (also known as masking) can be studied by using an ultradian 7-h light/dark (3.5:3.5 LD) cycle that disrupts the oscillator<sup>6</sup>. Under the ultradian regime, wild types confined their activity mostly to the dark portion of the ultradian cycle ( $84 \pm 4\%$ ). In contrast, the *Opn4<sup>aDTA/aDTA</sup>* animals were active nearly randomly across an ultradian day, irrespective of light ( $64 \pm 3\%$ ; Fig. 4d). This result demonstrates that melanopsin cells are essential for a direct light-driven physiological response that is independent of the circadian oscillator.

We have shown that the loss of ipRGCs does not influence image formation and therefore the involvement of this ganglion cell class in classical vision is only modulatory<sup>13,23</sup>. In contrast, irradiance-dependent NIF functions are substantially impaired in the absence of ipRGCs. Given that the ipRGCs constitute less than 2% of the total RGC population, it is striking that light information for circadian photoentrainment and PLR is conveyed predominantly through these cells. Moreover, the ability to form pattern vision does not affect photoentrainment.

## METHODS SUMMARY

**Animals.** All experiments were conducted in accordance with National Institutes of Health guidelines and were approved by institutional animal care and use committees of the universities involved.

**Behavioural analyses.** We used behavioural tests measuring the integrity of the outer retina (electroretinograms)<sup>23</sup>, eye-tracking functions (optokinetic nystagmus)<sup>24</sup>, visual acuity (optomotor)<sup>25</sup>, object identification (water maze)<sup>26</sup>, pupil constriction (PLR)<sup>8,12</sup>, the period of the circadian oscillator (wheel running activity)<sup>10,11</sup>, the adjustment of the circadian clock to different light stimulations (circadian photoentrainment, phase shifting, and constant light)<sup>10,11</sup> and direct light effects on activity (masking)<sup>9</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 July 2007; accepted 7 February 2008.

Published online 23 April 2008.

- Hubel, D. H. *Eye, Brain, and Vision* (Freeman, New York, 1988).
- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Freedman, M. S. *et al.* Regulation of mammalian circadian behavior by non-rod, non-cone, ocular photoreceptors. *Science* **284**, 502–504 (1999).
- Czeisler, C. A. *et al.* Suppression of melatonin secretion in some blind patients by exposure to bright light. *N. Engl. J. Med.* **332**, 6–11 (1995).
- Hattar, S. *et al.* Melanopsin and rod–cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
- Panda, S. *et al.* Melanopsin is required for non-image-forming photic responses in blind mice. *Science* **301**, 525–527 (2003).
- Lucas, R. J. *et al.* Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).

- Mrosovsky, N. & Hattar, S. Impaired masking responses to light in melanopsin-knockout mice. *Chronobiol. Int.* **20**, 989–999 (2003).
- Panda, S. *et al.* Melanopsin (*Opn4*) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
- Ruby, N. F. *et al.* Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
- Lucas, R. J., Douglas, R. H. & Foster, R. G. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neurosci.* **4**, 621–626 (2001).
- Dacey, D. M. *et al.* Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature* **433**, 749–754 (2005).
- Wong, K. Y., Dunn, F. A., Graham, D. M. & Berson, D. M. Synaptic influences on rat ganglion-cell photoreceptors. *J. Physiol. (Lond.)* **582**, 279–296 (2007).
- Hattar, S. *et al.* Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J. Comp. Neurol.* **497**, 326–349 (2006).
- Gooley, J. J., Lu, J., Fischer, D. & Saper, C. B. A broad role for melanopsin in nonvisual photoreception. *J. Neurosci.* **23**, 7093–7106 (2003).
- Hannibal, J. & Fahrenkrug, J. Target areas innervated by PACAP-immunoreactive retinal ganglion cells. *Cell Tissue Res.* **316**, 99–113 (2004).
- Morin, L. P., Blanchard, J. H. & Provencio, I. Retinal ganglion cell projections to the hamster suprachiasmatic nucleus, intergeniculate leaflet, and visual midbrain: bifurcation and melanopsin immunoreactivity. *J. Comp. Neurol.* **465**, 401–416 (2003).
- Sollars, P. J. *et al.* Melanopsin and non-melanopsin expressing retinal ganglion cells innervate the hypothalamic suprachiasmatic nucleus. *Vis. Neurosci.* **20**, 601–610 (2003).
- Maxwell, F., Maxwell, I. H. & Glode, L. M. Cloning, sequence determination, and expression in transfected cells of the coding sequence for the tox 176 attenuated diphtheria toxin A chain. *Mol. Cell. Biol.* **7**, 1576–1579 (1987).
- Wee, R., Castrucci, A. M., Provencio, I., Gan, L. & Van Gelder, R. N. Loss of photic entrainment and altered free-running circadian rhythms in *math5<sup>-/-</sup>* mice. *J. Neurosci.* **22**, 10427–10433 (2002).
- Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
- Barnard, A. R., Hattar, S., Hankins, M. W. & Lucas, R. J. Melanopsin regulates visual processing in the mouse retina. *Curr. Biol.* **16**, 389–395 (2006).
- Iwakabe, H., Katsura, G., Ishibashi, C. & Nakanishi, S. Impairment of pupillary responses and optokinetic nystagmus in the mGluR6-deficient mouse. *Neuropharmacology* **36**, 135–143 (1997).
- Prusky, G. T., Alam, N. M., Beekman, S. & Douglas, R. M. Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system. *Invest. Ophthalmol. Vis. Sci.* **45**, 4611–4616 (2004).
- Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nature Protocols* **1**, 848–858 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Mackes and G. Harrison for help in genotyping the animals; R. Kuruvilla, M. Van Doren, B. Wendland, M. Halpern, M. Caterina, C.-Y. Su, J. Bradley and laboratory members in the Biology Department at the Johns Hopkins University for scientific discussions and comments on the manuscript. This work was supported by grants from the National Institutes of Health (to S. Hattar and K.-W.Y.), the Biotechnology and Biological Sciences Research Council (to R.J.L.) and the David and Lucile Packard and Alfred P. Sloan Foundations (to S. Hattar).

**Author Contributions** A.D.G. and S. Hattar wrote the paper. J.L.E., R.J.L., D.M.B. and T.C.B. gave helpful comments on the manuscript. A.D.G., J.L.E. and C.M.A. in S. Hattar's laboratory performed all the behavioural studies on the *aDTA* homozygous animals, as well as the X-gal staining of the *Opn4<sup>aDTA/tau-LacZ</sup>* and *Opn4<sup>tau-LacZ/+</sup>* animals, the morphology of the retina, the cholera toxin injections, the water maze and the optomotor studies. D.M.B. helped in analysing the brains of the *Opn4<sup>aDTA/tau-LacZ</sup>* and the cholera-toxin-injected animals. G.S.L. and A.R.B. in R.J.L.'s laboratory conducted all the behavioural studies on the *aDTA* heterozygous animals, and the electroretinogram studies. T.C.B. provided the construct and suggestions for the *aDTA* targeting strategy. H.C. made the optokinetic nystagmus recordings. H.-W.L. in K.-W.Y.'s laboratory performed the melanopsin immunostaining on *aDTA* heterozygous mice. Animals were first conceived in K.-W.Y.'s laboratory and produced by S. Hattar and S. Haq to the chimeric stage. Germline transmission was obtained independently in the laboratories of S. Hattar (with help from H.Z.) and K.-W.Y. All other authors helped in the planning, technical support and discussions of experiments.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S. Hattar ([shattar@jhu.edu](mailto:shattar@jhu.edu)) or R.J.L. ([robert.lucas@manchester.ac.uk](mailto:robert.lucas@manchester.ac.uk)).

## METHODS

**aDTA mice.** Using the homologous arms that we used previously<sup>3</sup>, we targeted the *aDTA* gene to the melanopsin locus. The targeting construct contained the diphtheria toxin A subunit and the neomycin resistance genes. The construct was flanked by 4.4 kilobases (kb) 5' of the ATG site of the mouse melanopsin gene, and a 1.6-kb fragment containing 654 base pairs (bp) of exon 9 plus a 946-bp 3' untranslated region. After electroporation of the linearized construct into 129.1 mouse strain embryonic stem cells and drug selection (400  $\mu\text{g ml}^{-1}$  G418), one positive embryonic stem clone was injected into C57BL/6 blastocysts. Chimaeric animals were mated to C57BL/6 mice to produce heterozygous animals.

**Immunostaining.** Whole retinas from *Opn4<sup>aDTA/+</sup>* and wild-type animals, fixed in 70% ethanol, were immunostained with the C terminus melanopsin antibody (1:500 dilution). Fluorescently conjugated secondary antibody (Alexa Fluor 488 goat anti-rabbit IgG (1:1,000 dilution; Molecular Probes) was applied for 1–3 h. **Staining with X-gal.** Animals anaesthetized by intraperitoneal injection of Avertin (20 ml  $\text{kg}^{-1}$ ) were perfused intracardially with 4% paraformaldehyde, and brains and eyes were isolated. Eye-cups or brain sections (50  $\mu\text{m}$ ) were first incubated in buffer B (100 mM phosphate buffer at pH 7.4, 2 mM  $\text{MgCl}_2$ , 0.01% sodium deoxycholate, 0.02% IGEPAL) then stained for 3 days in buffer B plus 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide and 1 mg  $\text{ml}^{-1}$  X-gal.

**Cholera toxin injections in the eye.** Mice were anaesthetized with Ketamine (80 mg  $\text{kg}^{-1}$ )/Xylazine (8 mg  $\text{kg}^{-1}$ ). Eyes were injected intravitreally with 2  $\mu\text{l}$  of cholera toxin B subunit conjugated with Alexa Fluor 488 or Alexa Fluor 555 (Invitrogen). Three days after injection, brains were isolated and sectioned.

**Electroretinograms.** Animals maintained under 12:12 light/dark cycles for three days were used to collect data within the light phase (after 50 min of dark adaptation). The electroretinogram set-up was similar to that used previously<sup>23</sup>. Mydriatics (1% tropicamide and 2.5% phenylephrine) and hypromellose solution (0.3%) were used to dilate the pupil and retain corneal moisture in anaesthetized mice. A signal conditioner (Model 1902 Mark III; Cambridge Electronic Design) differentially amplified ( $\times 3,000$ ) and filtered (bandpass filter cutoff 0.5–200 Hz) the signal before it was digitized (Model 1401 digitizer; Cambridge Electronic Design) and recorded (sampling rate 10 kHz) by means of Signal 2.15 software (Cambridge Electronic Design).

White light (10 ms in duration) was provided by a xenon arc source (Cairn Research). Neutral density filters (Edmund Optics) were used (unattenuated intensity 320  $\mu\text{W cm}^{-2}$  at the cornea). For scotopic recordings, flashes were administered in the dark and testing began with the dimmest stimulus. Depending on the intensity, stimuli were presented at a rate of 0.5–0.2 Hz, and 6–20 repetitions were collected and averaged. Photopic electroretinograms were recorded by presentation of an unattenuated light (20 stimuli presented at a rate of 1 Hz) against blue-filtered (Grass blue filter; Astro-Med) rod-saturating background light (160  $\mu\text{W cm}^{-2}$ ).

**Optokinetic nystagmus responses.** A mouse stabilized with a head post was placed into an acrylic holder in a 12-inch diameter drum. Computer-generated stimuli were projected down onto the drum walls. Black and white stripes 4° in width were rotated at 5°  $\text{s}^{-1}$ . Eye movements of mice were captured with an infrared video system (ISCAN). The fast saccade components were counted in 30-s bins with an algorithm that evaluated the high-velocity eye movements that followed a low-velocity movement in the opposite direction.

**Visual acuity.** A virtual cylinder OptoMotry (Cerebral Mechanics) was used to determine visual acuity by measuring the image-tracking reflex of mice<sup>25</sup>. A sine-wave grating was projected on the screen rotating in a virtual cylinder. The animal was assessed for a tracking response on stimulation for about 5 s. All acuity thresholds were determined by using the staircase method with 100% contrast.

**Morris water maze.** To assess the ability of mice to detect a visual cue, we trained the animals to find a platform marked by a 10-cm tall, high-contrast visual cue under bright light (500 lx) in an 85-cm pool. On day 1, mice were trained with four trials 15 min apart. On the following day, latency to find the island was recorded first with the cue and then without it.

**PLR.** All animals were dark-adapted for at least 1 h, and the eye of each animal receiving the photic stimulus was treated with 0.1% atropine before the start of recording. Measurements were restricted to the middle of the subjective day (CT4–8). One eye of each mouse was digitally captured at a frequency of one image per second for 63 s with a charge-coupled-device camera. The light stimuli (xenon arc light source) consisted of a 60-s pulse at an intensity of 3.8  $\text{mW cm}^{-2}$  or 1.8  $\mu\text{W cm}^{-2}$  of white light.

For the *Opn4<sup>aDTA/aDTA</sup>* experiments, the eye receiving the photic stimulus was treated with 1% atropine. While one eye received light stimulation from a 470-nm light-emitting-diode light source (E27-B24; 161  $\mu\text{W cm}^{-2}$ ; Super Bright LEDs), a digital camcorder (DCRHC96; Sony) was used to record from the other eye (for 30 s) at 30 frames  $\text{s}^{-1}$  under a 940-nm light (LDP). The digital video recoding was deconstructed to individual frames with Blaze MediaPro software (Mystik Media). The percentage pupil constriction was calculated as the percentage of pupil area at 30 s after initiation of the stimulus (steady state) relative to the dilated pupil size.

**Wheel running activity.** Mice were placed in cages with a 4.5-inch running wheel, and their activity was monitored with VitalView software (Mini Mitter). The period was calculated with ClockLab (Actimetrics). For phase-shifting experiments, each animal was exposed to a light pulse (1,500 lx; CT16) for 15 min. After 41 days of constant dark, mice were re-entrained to 12:12 light/dark cycles for 19 days. Animals were then exposed to two jet-lag models: 16 days of a 6-h advance followed by 32 days of a 6-h delay. During the last two weeks of this treatment the animals were tested for PLR. Animals were then exposed to constant light for three weeks followed by ultradian 3.5:3.5 light/dark cycles.

## LETTERS

# The aryl hydrocarbon receptor links T<sub>H</sub>17-cell-mediated autoimmunity to environmental toxins

Marc Veldhoen<sup>1</sup>, Keiji Hirota<sup>1</sup>, Astrid M. Westendorf<sup>2,3</sup>, Jan Buer<sup>2</sup>, Laure Dumoutier<sup>4</sup>, Jean-Christophe Renaud<sup>4</sup> & Brigitta Stockinger<sup>1</sup>

The aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor best known for mediating the toxicity of dioxin<sup>1</sup>. Environmental factors are believed to contribute to the increased prevalence of autoimmune diseases, many of which are due to the activity of T<sub>H</sub>17 T cells, a new helper T-cell subset characterized by the production of the cytokine IL-17. Here we show that in the CD4<sup>+</sup> T-cell lineage of mice AHR expression is restricted to the T<sub>H</sub>17 cell subset and its ligation results in the production of the T<sub>H</sub>17 cytokine interleukin (IL)-22. AHR is also expressed in human T<sub>H</sub>17 cells. Activation of AHR by a high-affinity ligand during T<sub>H</sub>17 cell development markedly increases the proportion of T<sub>H</sub>17 T cells and their production of cytokines. CD4<sup>+</sup> T cells from AHR-deficient mice can develop T<sub>H</sub>17 cell responses, but when confronted with AHR ligand fail to produce IL-22 and do not show enhanced T<sub>H</sub>17 cell development. AHR activation during induction of experimental autoimmune encephalomyelitis causes accelerated onset and increased pathology in wild-type mice, but not AHR-deficient mice. AHR ligands may therefore represent co-factors in the development of autoimmune diseases.

The AHR is a ligand-dependent transcription factor that mediates a range of critical cellular events in response to halogenated aromatic hydrocarbons and non-halogenated polycyclic aromatic hydrocarbons such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)<sup>1</sup>. AHR expression is ubiquitous in vertebrate cells, suggesting important and widespread roles, but the physiological role of AHR is not yet understood<sup>2</sup>. Mice with a targeted mutation of the *Ahr* gene provided unequivocal evidence that the AHR is crucial to TCDD-induced toxicity and suggested a function for AHR in liver development<sup>3–5</sup>.

Gene array analysis of CD4<sup>+</sup> effector T-cell subsets showed that the T<sub>H</sub>17 CD4<sup>+</sup> T-cell subset, in addition to the lineage-defining transcription factor *Rorc* (encoding ROR- $\gamma$ t, an isoform of ROR- $\gamma$  expressed in the thymus)<sup>6</sup>, expresses *Ahr*. Quantitative polymerase chain reaction (PCR) analysis of CD4<sup>+</sup> effector T-cell subsets from wild-type and AHR-deficient C57BL/6 (B6) mice established that the lineage-defining transcription factors *T-bet*, *Gata3* and *Rorc* are expressed in a comparable manner as are the marker cytokines interferon (IFN)- $\gamma$ , IL-4, IL-17A and IL-17F (Fig. 1a). There is also similar *Foxp3* expression in CD4<sup>+</sup> T cells from B6 and AHR-deficient mice (Supplementary Fig. 1b). AHR was only induced under T<sub>H</sub>17-cell-inducing conditions (IL-6 and transforming growth factor (TGF)- $\beta$ ) with levels of expression similar to liver (Fig. 1a, d). Under our conditions we found no AHR expression in cultures with TGF- $\beta$  alone (induced T-regulatory cells; iT<sub>reg</sub>) or IL-6 alone and no expression in natural T<sub>reg</sub> cells (Supplementary Fig. 1a). Notably, AHR expression is also found in human T<sub>H</sub>17 cells (Fig. 1b).

Although CD4<sup>+</sup> T cells from AHR-deficient mice could differentiate into T<sub>H</sub>17 cells, they lacked the expression of IL-22 (Fig. 1c). IL-22, originally defined as a hepatocyte-stimulating factor<sup>7</sup>, is co-expressed with IL-17 by T<sub>H</sub>17 T cells and its expression is thought to be enhanced by dendritic-cell-derived IL-23 (refs 8, 9). The biological functions of IL-22 are not fully understood: on the one hand IL-22 seems to be pro-inflammatory, inducing dermal inflammation and pro-inflammatory gene expression in the skin<sup>8,10</sup>, on the other hand IL-22 delivery ameliorates T-cell-mediated liver injury in T-cell-mediated hepatitis<sup>11,12</sup>.

AHR resides in the cytoplasm in complex with Hsp90 until binding of ligand triggers conformational changes resulting in an exchange of Hsp90 for the nuclear translocation component ARNT (reviewed in ref. 1). *Arnt* was found expressed in all CD4<sup>+</sup> T-cell subsets (data not shown). Ligation of AHR by 6-formylindolo[3,2-b]carbazole (FICZ), a tryptophan-derived photoproduct that is thought to be an endogenous ligand with high affinity for the AHR receptor, upregulates genes encoding xenobiotic metabolizing cytochrome P450 enzymes such as *Cyp1a1* (ref. 13).

To test whether exposure of T cells to FICZ influences differentiation of naive CD4<sup>+</sup> T cells to effector cells, we added FICZ during the *in vitro* differentiation of CD4<sup>+</sup> effector T-cell subsets. The addition of FICZ did not induce *Ahr* expression or its downstream target *Cyp1a1* in T<sub>H</sub>0, T<sub>H</sub>1, T<sub>H</sub>2 and iT<sub>reg</sub> T-cell subsets (Fig. 2a) and did not alter their expression of *Ifng*, *Il4*, *Il21* nor their lineage-defining transcription factors (Supplementary Fig. 1c). However, the presence of FICZ during T<sub>H</sub>17-cell-inducing conditions led to strong upregulation of *Il17a*, *Il17f* and particularly of *Il22* mRNA expression (Fig. 2a and Supplementary Fig. 2a).

Comparison of T<sub>H</sub>17 differentiation in CD4<sup>+</sup> T cells from wild-type B6 and AHR-deficient mice by intracellular staining showed that exposure of B6 CD4<sup>+</sup> T cells to FICZ under T<sub>H</sub>17-cell-inducing conditions strongly enhanced IL-17A and IL-17F production and increased the proportion and staining intensity of cells producing IL-22 (Fig. 2b, top panels and Supplementary Fig. 2b). In contrast, IL-17A and IL-17F production was attenuated in T<sub>H</sub>17 cells from AHR-deficient mice and no IL-22 was detectable whether FICZ was present or not (Fig. 2b, bottom panels). A similar response was seen with another AHR ligand,  $\beta$ -naphthoflavone, which has lower affinity for AHR and consequently requires about a 10-fold higher dose compared with FICZ (Supplementary Fig. 2c). Similar to mouse, human T<sub>H</sub>17 cells reacted to AHR ligation with increased expression of *IL17A*, *IL17F* and *IL22* as well as induction of *CYP1A1* (Fig. 2c).

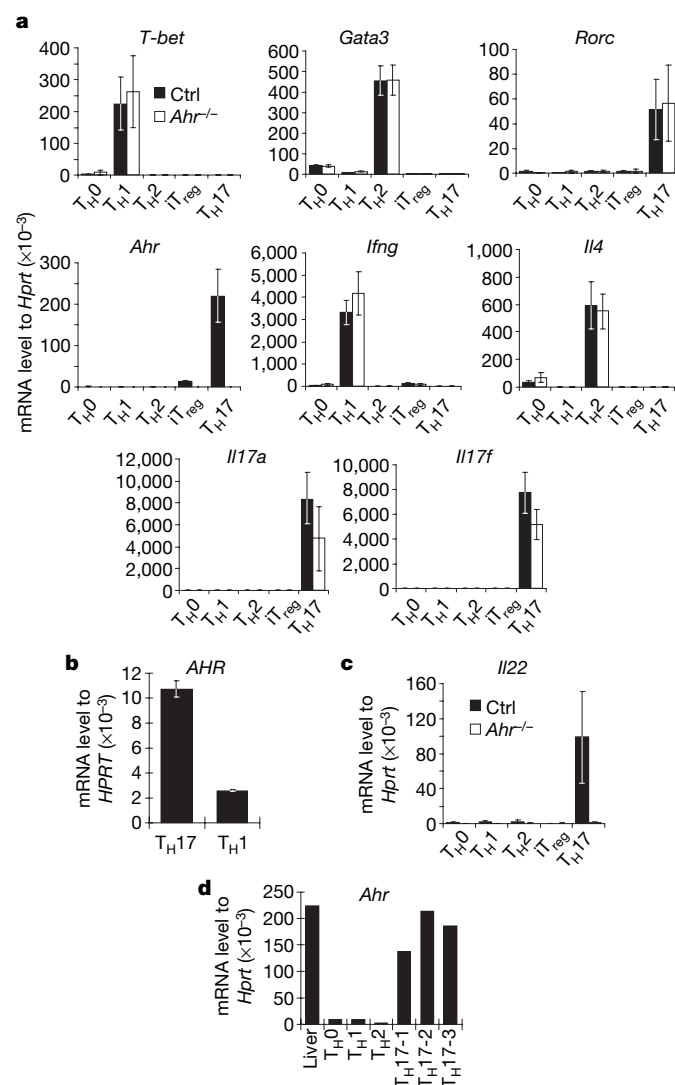
To test whether *Ahr* expression on its own is essential and sufficient to drive IL-22 expression, we performed retroviral transduction of sorted naive AHR-deficient CD4<sup>+</sup> T cells with an AHR-green

<sup>1</sup>Division of Molecular Immunology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW71AA, UK. <sup>2</sup>Institute for Medical Microbiology, University Hospital Essen, D-45122, Germany. <sup>3</sup>Helmholtz Center for Infection Research, D-38124 Braunschweig, Germany. <sup>4</sup>Ludwig Institute for Cancer Research, Brussels branch, and Experimental Medicine Unit, Université Catholique de Louvain, B-1200 Brussels, Belgium.



fluorescent protein (GFP) construct or a GFP vector control construct. Transduction under neutral,  $T_H1$ ,  $T_H2$  or  $iT_{reg}$  conditions did not reconstitute IL-22 expression even in the presence of FICZ (Supplementary Fig. 3). However, under  $T_H17$ -cell-inducing conditions, reconstitution of AHR expression by retroviral transduction induced expression of IL-22 (Fig. 3a, top-right panel) and increased the proportion of IL-17-producing cells (Fig. 3b, top-right panel). Exposure to FICZ resulted in a substantial increase in IL-22 (Fig. 3a, bottom-right panel) as well as the enhanced expression of IL-17A (Fig. 3b, bottom right panel). PCR with reverse transcription (RT-PCR) analysis of cultured *Ahr*-transduced  $T_H17$  cells confirmed the increase in *Il17* expression, the enhancement of *Il22* as well as the induction of the *Ahr* target *Cyp11a1* on exposure to FICZ (Fig. 3c). Thus, *Ahr* expression is only functional in  $CD4^+$  T cells that have differentiated to the  $T_H17$  lineage.

Next we immunized B6 mice with myelin oligodendrocyte peptide 35–55 (MOG<sub>35–55</sub>) in complete Freund's adjuvant (CFA), removed draining lymph nodes 7 days later and isolated  $CCR6^+CD4^+$  T cells (which are enriched in  $T_H17$  cells<sup>14,15</sup>) to test whether *Ahr* is also



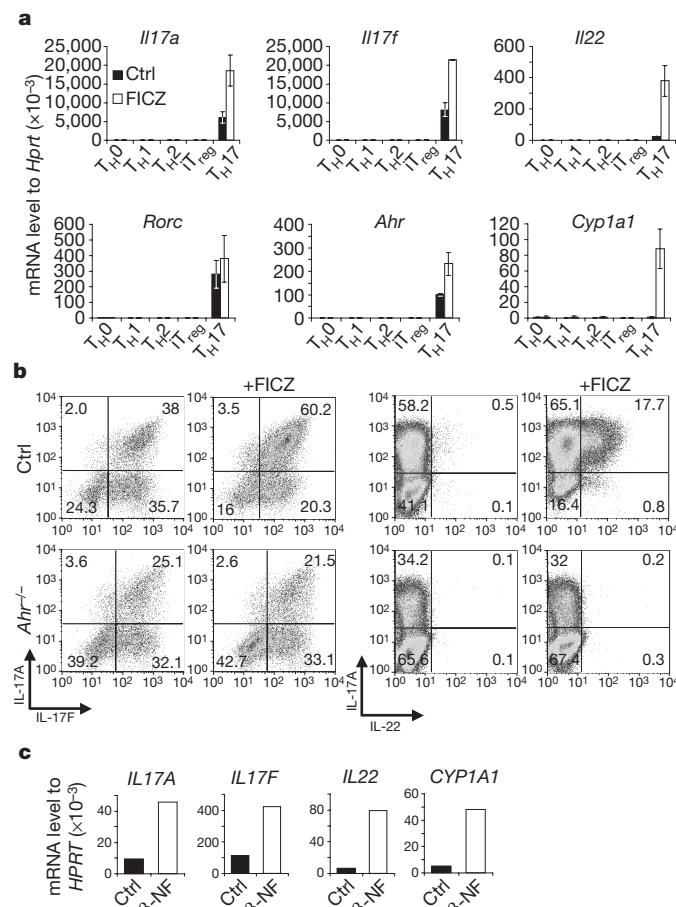
**Figure 1 | AHR is selectively expressed in the  $T_H17$  cell subset.**

**a**, Fluorescence-activated cell sorter (FACS)-sorted naive  $CD4^+$  T cells from B6 or *Ahr*-deficient mice stimulated under  $T_H0$ ,  $T_H1$ ,  $T_H2$ ,  $iT_{reg}$  or  $T_H17$  conditions and harvested on day 5 for quantitative PCR. mRNA levels, normalized to *Hprt* expression  $\pm$  s.d., are shown. *Rorc*, mRNA encoding ROR- $\gamma$ t. **b**, mRNA levels  $\pm$  s.d. of *AHR* in human  $CD4^+$  T cells stimulated under  $T_H1$  or  $T_H17$  conditions. **c**, mRNA levels  $\pm$  s.d. for *Il22* in  $CD4^+$  T cell subsets from B6 or *Ahr*-deficient mice. **d**, mRNA levels of *Ahr* in  $CD4^+$  T-cell subsets and  $T_H17$  cells from three mice compared with mouse liver.

expressed in  $T_H17$  cells generated *in vivo*.  $CCR6^+CD4^+$  T cells expressed *Ahr*, *Il17a*, *Il17f*, *Il22* and *Rorc*, whereas the  $CCR6^-CD4^+$  T-cell fraction lacked expression of *Ahr* and  $T_H17$  markers, confirming that physiological differentiation *in vivo* recapitulates *in vitro* differentiation (Fig. 4a).

$T_H17$  T cells have a prominent role in the pathology of autoimmune diseases such as experimental autoimmune encephalomyelitis (EAE), which is induced by immunization with MOG peptide 35–55 and CFA. Analysis of spinal cord at the height of the EAE response on day 18 showed increased numbers of  $CD4^+$  T cells producing IL-17 and IL-22 in B6 mice treated with MOG<sub>35–55</sub>/CFA and FICZ, whereas *Ahr*-deficient mice showed reduced numbers of IL-17-producing T cells and minimal numbers of IL-22 producers (Fig. 4b). There was no difference in the numbers of *Foxp3*-expressing regulatory T cells in spinal cord in the three groups of mice.

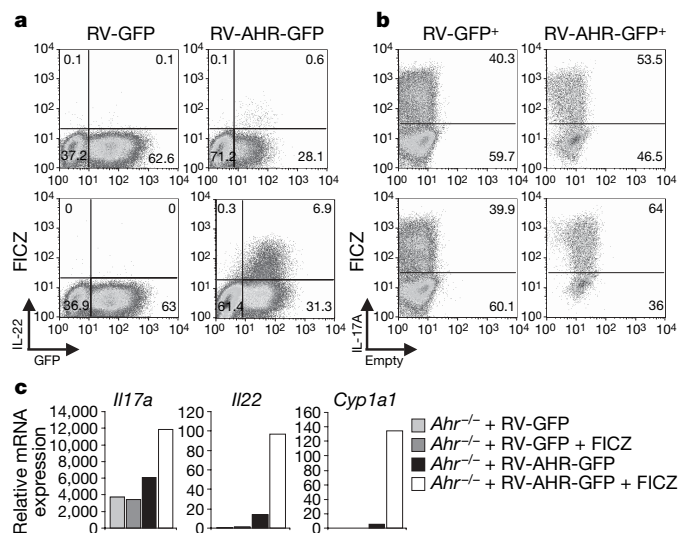
Whereas B6 mice developed EAE with a mean day of onset of 13.9, *Ahr*-deficient mice developed EAE with delayed kinetics (mean day of onset 15.6) in line with the attenuated  $T_H17$  differentiation seen *in vitro*. Despite the delayed onset of EAE, most *Ahr*-deficient mice succumbed to disease eventually (Fig. 4c, e). Stimulation of AHR by inclusion of FICZ in the antigen emulsion accelerated the onset (day 11.7) and increased the severity of EAE in B6 mice, but, as expected, did not influence the onset or severity of EAE in *Ahr*-deficient mice



**Figure 2 | AHR ligation promotes the  $T_H17$  cell programme.** **a**, Mean mRNA levels  $\pm$  s.d. from B6  $CD4^+$  effector subsets generated in the presence or absence of 200 nM FICZ. **b**,  $CD4^+$  T cells from B6 (top panels) and *Ahr*-deficient mice (bottom panels) cultured under  $T_H17$ -cell-inducing conditions in the presence or absence of 200 nM FICZ and stained for expression of IL-17A versus IL-17F (left panels) and IL-17A versus IL-22 (right panels) after re-stimulation with phorbol dibutyrate and ionomycin. A summary of all experiments is shown in Supplementary Fig. 2b. **c**, mRNA levels of *IL17A*, *IL17F*, *IL22* and *CYP11A1* in human  $CD4^+$  T cells stimulated under  $T_H17$ -cell-inducing conditions in the presence or absence of 3  $\mu$ M  $\beta$ -naphthoflavone ( $\beta$ -NF).

(day 15.8) (Fig. 4c, e). To assess the influence of AHR deficiency in haematopoietic or non-haematopoietic cells, we also induced EAE in chimaeras constructed either by injection of AHR-deficient bone marrow into irradiated wild-type mice (AHR-B6) or by injection of B6 wild-type bone marrow into irradiated AHR-deficient mice (B6-AHR). AHR-B6 chimaeras showed attenuated EAE like AHR-deficient mice, whereas B6-AHR chimaeras developed EAE with kinetics and severity similar to wild-type mice (Fig. 4d).

Studies on the effect of AHR stimulation on the immune system have so far focused exclusively on TCDD as a ligand because of its toxicological relevance. Although adverse effects of TCDD on immune responses are well documented<sup>16</sup>, no direct measurements of AHR expression on highly purified polarized subsets of CD4<sup>+</sup> T cells have been reported. TCDD, which cannot be metabolized and therefore causes prolonged stimulation of AHR in many cells of the body, is known to induce profound suppression of immune responses in wild-type, but not AHR-deficient, mice<sup>16</sup>, but despite decades of research the underlying mechanisms for this profound toxicity remain unclear. More recently, it has been suggested that TCDD promotes the generation of regulatory T cells<sup>17</sup> while causing the premature decline of activated CD4<sup>+</sup> T cells. In our hands, stimulation of AHR by FICZ did not influence the number of regulatory T cells during MOG<sub>35–55</sub>-induced EAE responses (see Fig. 4b). It is conceivable that the high toxicity of TCDD for effector cells causes a proportional shift in regulatory T cells—which are, in general, more resistant to many depletion regimes<sup>18,19</sup>—due to the death of other cells, rather than an actual expansion in numbers. Although TCDD may be the classical AHR ligand used in the toxicology field to analyse the effects of AHR activation, it is clear that it is not a natural ligand and it is widely recognized that pollutants such as TCDD are unlikely to have provided the evolutionary pressure for the function of this highly conserved system. There is ongoing debate in the toxicology field about what is the most relevant physiological ligand for AHR<sup>20</sup>. Nevertheless, compelling indirect evidence shows that ultraviolet photoproducts of tryptophan, such as the high-affinity ligand FICZ studied in our paper, could be synthesized *in vivo* as exposure of human skin to ultraviolet light induces CYP1A1 (ref. 21). Future



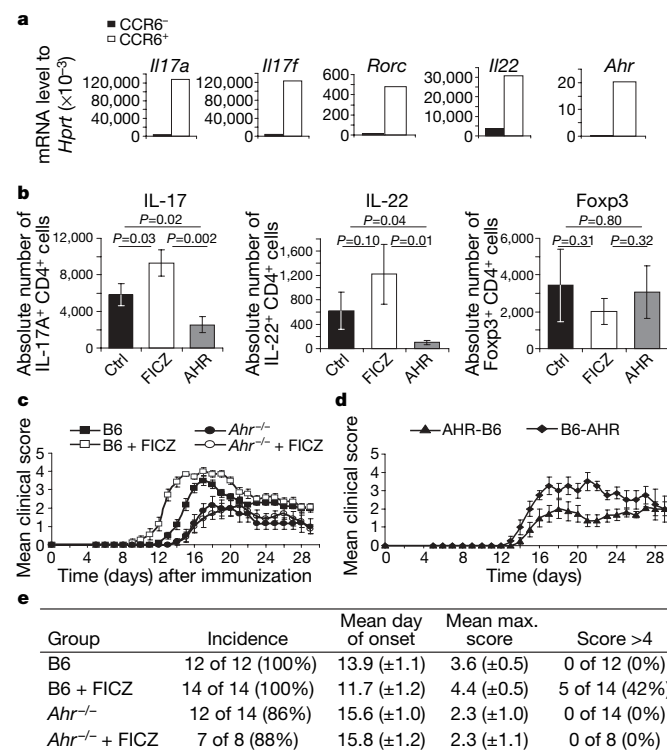
**Figure 3 | Retroviral transduction of AHR restores IL-22 expression.**

**a, b**, FACS-sorted naive CD4<sup>+</sup> T cells from AHR-deficient mice cultured under T<sub>H</sub>17-cell-inducing conditions and transduced with vector control (RV-GFP) or AHR-containing construct (RV-AHR-GFP) in the presence (bottom panels) or absence (top panels) of FICZ. IL-22 intracellular staining versus GFP expression (**a**) and IL-17A expression in gated GFP<sup>+</sup> cells (**b**) were assessed. **c**, Quantitative PCR for *Il17a*, *Il22* and *Cyp1a1* in T<sub>H</sub>17 cells from AHR-deficient mice transduced with control retroviral vector, control vector in the presence of 200 nM FICZ, AHR-containing vector, or AHR-containing vector in the presence of FICZ.

studies focusing on additional physiological ligands of AHR that can be metabolized by the CYP1 enzymes and therefore cause only transient AHR activation may give further insights into the consequences of AHR stimulation in T<sub>H</sub>17 cells.

Our data show that AHR, in addition to promoting the expression of IL-22, enhances T<sub>H</sub>17 cell development and the expression levels of IL-17A and IL-17F, and consequently increases autoimmune pathology. Blockade of IL-17A with an auto-vaccine as well as neutralizing antibody can completely prevent the development of EAE<sup>22,23</sup>, emphasizing the central role of IL-17A in the pathogenesis of this autoimmune disease. Although it has been suggested that IL-22 may contribute to autoimmune pathology because, like IL-17, it disrupts blood–brain barrier tight junctions<sup>24</sup>, IL-22-deficient mice do not seem to have altered susceptibility to EAE induction<sup>25</sup>, suggesting that this cytokine may be dispensable for the development of pathology in EAE. Thus, the enhancement of IL-17 production by AHR ligation may be more crucial than the induction of IL-22 in determining disease severity.

It is currently thought that IL-22 expression depends on induction by IL-23 (refs 8, 9). Re-stimulation of lymph node cells from mice immunized with MOG<sub>35–55</sub>/CFA or MOG<sub>35–55</sub>/CFA and FICZ in the presence or absence of IL-23 showed that IL-23 increased the proportion of IL-17-producing cells both from wild-type and AHR-deficient mice (Supplementary Fig. 4, left), but only had a small effect on the production of IL-22 (Supplementary Fig. 4b, right), suggesting that IL-23 has some effect independent of AHR ligation, although clearly stimulation of AHR seems to be the dominant pathway for IL-22 production.



**Figure 4 | EAE is enhanced by AHR ligation.** **a**, RT-PCR analysis for *Il17a*, *Il17f*, *Il22*, *Rorc* (ROR-γt) and *Ahr* in FACS-sorted CCR6<sup>+</sup> CD4<sup>+</sup> T cells or CCR6<sup>+</sup> CD4<sup>+</sup> T cells from draining lymph nodes 7 days after MOG<sub>35–55</sub> immunization. **b**, Mean numbers of IL-17A<sup>+</sup>, IL-22<sup>+</sup> and Foxp3<sup>+</sup> cells ± s.d. in spinal cord 18 days after immunization of B6 mice (*n* = 4) with MOG<sub>35–55</sub> (Ctrl) or MOG<sub>35–55</sub> + FICZ (FICZ) and of AHR-deficient mice with MOG<sub>35–55</sub> (AHR) (*n* = 4). **c**, Mean clinical EAE scores ± s.e.m. of B6 (*n* = 12) or AHR-deficient mice (*n* = 12) immunized with MOG<sub>35–55</sub> in the absence or presence of 600 ng FICZ. **d**, Mean clinical EAE scores ± s.e.m. of MOG<sub>35–55</sub>-immunized chimaeras from AHR-deficient donors into irradiated B6 hosts (AHR-B6 *n* = 8), or bone marrow from B6 donors into irradiated AHR-deficient hosts (B6-AHR *n* = 8). **e**, Incidence, mean day of onset and mean maximal scores for the mice in **c**. P values are *P* = 0.0013 for mean day of onset in B6 versus B6 + FICZ and *P* = 0.03 for B6 versus AHR-deficient.

How AHR interacts with the  $T_H17$  pathway is currently unknown, but there is a substantial amount of literature (reviewed in ref. 26) describing interactions of AHR with other key regulatory proteins including nuclear factor- $\kappa B$ <sup>27</sup>, which has a role in the induction of EAE<sup>28</sup>. The core nucleotide sequence to which the nuclear AHR complex binds, also termed the 'xenobiotic responsive element'<sup>29</sup>, occurs frequently in the mammalian genome and is also represented in IL-17A, IL-17F, IL-22 and ROR- $\gamma t$ . Although basal expression of AHR and IL-22 is detectable in  $T_H17$  T cells in the apparent absence of a ligand, there are numerous endogenous agents that can activate AHR, such as prostaglandins, bilirubin at high concentration, modified low-density lipoprotein and various modifications of tryptophan, whose ultraviolet-light-irradiated photoconversion into the high-affinity ligand FICZ is only one example (reviewed in ref. 20).

Autoimmune diseases are multifactorial, depending on intrinsic components such as genetics, hormones or age, and environmental factors, including infections, diet, drugs and chemicals. The increasing prevalence of certain autoimmune diseases in highly industrialized countries is probably connected to such environmental factors.

Our data linking a transcription factor responsive to environmental pollutants to the  $T_H17$  programme open intriguing possibilities regarding the potential of such factors to initiate or augment autoimmune conditions, and warrant closer examination of a possible role of AHR in human autoimmune diseases.

## METHODS SUMMARY

**Mice.** C57BL/6 (B6) and AHR-deficient mice on a B6 background (B6 BRA AHRKO)<sup>3</sup>, originally obtained from the Jackson Laboratory via A. Smith, were bred in the specified pathogen free (SPF) facility at NIMR. All animal experiments were done according to institutional guidelines and Home Office regulations.

**Human T-cell culture.** Human peripheral blood mononuclear cells from a healthy volunteer were isolated by Ficoll/Paque, and  $CD4^+$  T cells were purified by magnetic sorting and cultured at  $1.5 \times 10^5$  cells per well in plates coated with anti-CD3 ( $1 \mu g ml^{-1}$ ) and anti-CD28 ( $1 \mu g ml^{-1}$ ) in the presence of  $10 ng ml^{-1}$  IL-1 and  $40 ng ml^{-1}$  IL-6 ( $T_H17$  condition) or  $3 ng ml^{-1}$  IL-12,  $10 ng ml^{-1}$  IL-2 and  $10 \mu g ml^{-1}$  anti-TGF- $\beta$  ( $T_H1$  condition). Quantitative PCR analysis for  $T_H17$  markers and AHR expression was performed on day 4.

**In vitro T-cell differentiation and cytokine staining.** A detailed description of procedures is given in Methods.

**Real-time PCR.** The expression of mRNA for transcription factors and cytokines in  $CD4^+$  T-cell subsets was analysed 4–5 days after T-cell activation using specific primers from Applied Biosystems and expression was normalized to the house-keeping gene *Hprt*. More details and a list of primers used can be found in Methods.

**Retroviral transduction.** AHR was cloned into vector pIRES2-EGFP (Clontech) generating a bicistronic mRNA encoding AHR and, separated by an IRES element, EGFP. Viruses were generated by simultaneous  $CaCl_2$ -mediated transient transfections of 293T cells with three plasmids providing vector, Gag-Pol, and Env functions. Details of the transduction protocol are given in Methods.

**EAE induction.** EAE was induced and scored as described previously<sup>30</sup>. Some mice received 600 ng FICZ in the antigen emulsion. Draining lymph nodes were isolated 7 days after immunization. Spinal cord was isolated on day 18 after EAE induction for determination of cell numbers.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 November 2007; accepted 4 March 2008.

Published online 23 March 2008.

- Schmidt, J. V. & Bradfield, C. A. Ah receptor signaling pathways. *Annu. Rev. Cell Dev. Biol.* **12**, 55–89 (1996).
- Harper, P. A., Riddick, D. S. & Okey, A. B. Regulating the regulator: factors that control levels and activity of the aryl hydrocarbon receptor. *Biochem. Pharmacol.* **72**, 267–279 (2006).
- Schmidt, J. V., Su, G. H., Reddy, J. K., Simon, M. C. & Bradfield, C. A. Characterization of a murine AhR null allele: involvement of the Ah receptor in hepatic growth and development. *Proc. Natl Acad. Sci. USA* **93**, 6731–6736 (1996).
- Fernandez-Salguero, P. *et al.* Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor. *Science* **268**, 722–726 (1995).
- Mimura, J. *et al.* Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes Cells* **2**, 645–654 (1997).

- Ivanov, I. I. *et al.* The orphan nuclear receptor ROR $\gamma t$  directs the differentiation program of proinflammatory IL-17 $^+$  T helper cells. *Cell* **126**, 1121–1133 (2006).
- Dumoutier, L., Van Roost, E., Colau, D. & Renauld, J. C. Human interleukin-10-related T cell-derived inducible factor: molecular cloning and functional characterization as an hepatocyte-stimulating factor. *Proc. Natl Acad. Sci. USA* **97**, 10144–10149 (2000).
- Zheng, Y. *et al.* Interleukin-22, a  $T_H17$  cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **445**, 648–651 (2007).
- Liang, S. C. *et al.* Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. *J. Exp. Med.* **203**, 2271–2279 (2006).
- Boniface, K. *et al.* IL-22 inhibits epidermal differentiation and induces proinflammatory gene expression and migration of human keratinocytes. *J. Immunol.* **174**, 3695–3702 (2005).
- Radaeva, S., Sun, R., Pan, H. N., Hong, F. & Gao, B. Interleukin 22 (IL-22) plays a protective role in T cell-mediated murine hepatitis: IL-22 is a survival factor for hepatocytes via STAT3 activation. *Hepatology* **39**, 1332–1342 (2004).
- Zenewicz, L. A. *et al.* Interleukin-22 but not interleukin-17 provides protection to hepatocytes during acute liver inflammation. *Immunity* **27**, 647–659 (2007).
- Rannug, A. & Fritsche, E. The aryl hydrocarbon receptor and light. *Biol. Chem.* **387**, 1149–1157 (2006).
- Hirota, K. *et al.* Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. *J. Exp. Med.* **204**, 2803–2812 (2007).
- Acosta-Rodriguez, E. V. *et al.* Surface phenotype and antigenic specificity of human interleukin 17-producing T helper memory cells. *Nature Immunol.* **8**, 639–646 (2007).
- Kerkvliet, N. I. Recent advances in understanding the mechanisms of TCDD immunotoxicity. *Int. Immunopharmacol.* **2**, 277–291 (2002).
- Funatake, C. J., Marshall, N. B., Stepan, L. B., Mourich, D. V. & Kerkvliet, N. I. Cutting edge: activation of the aryl hydrocarbon receptor by 2,3,7,8-tetrachlorodibenzo-p-dioxin generates a population of  $CD4^+ CD25^+$  cells with characteristics of regulatory T cells. *J. Immunol.* **175**, 4184–4188 (2005).
- Minamimura, K., Gao, W. & Maki, T.  $CD4^+$  regulatory T cells are spared from deletion by antilymphocyte serum, a polyclonal anti-T cell antibody. *J. Immunol.* **176**, 4125–4132 (2006).
- Cox, A. L. *et al.* Lymphocyte homeostasis following therapeutic lymphocyte depletion in multiple sclerosis. *Eur. J. Immunol.* **35**, 3332–3342 (2005).
- Nguyen, L. P. & Bradfield, C. A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102–116 (2007).
- Katiyar, S. K., Matsui, M. S. & Mukhtar, H. Ultraviolet-B exposure of human skin induces cytokines P450 1A1 and 1B1. *J. Invest. Dermatol.* **114**, 328–333 (2000).
- Uytendhoe, C. & Van Snick, J. Development of an anti-IL-17A auto-vaccine that prevents experimental auto-immune encephalomyelitis. *Eur. J. Immunol.* **36**, 2868–2874 (2006).
- Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201**, 233–240 (2005).
- Kebir, H. *et al.* Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. *Nature Med.* **13**, 1173–1175 (2007).
- Kreyenborg, K. *et al.* IL-22 is expressed by Th17 cells in an IL-23-dependent fashion, but not required for the development of autoimmune encephalomyelitis. *J. Immunol.* **179**, 8098–8104 (2007).
- Okey, A. B. An aryl hydrocarbon receptor odyssey to the shores of toxicology: the Deichmann lecture, International Congress of Toxicology-XI. *Toxicol. Sci.* **98**, 5–38 (2007).
- Tian, Y., Rabson, A. B. & Gallo, M. A. Ah receptor and NF- $\kappa B$  interactions: mechanisms and physiological implications. *Chem. Biol. Interact.* **141**, 97–115 (2002).
- Hilliard, B., Samoilova, E. B., Liu, T. S., Rostami, A. & Chen, Y. Experimental autoimmune encephalomyelitis in NF- $\kappa B$ -deficient mice: roles of NF- $\kappa B$  in the activation and differentiation of autoreactive T cells. *J. Immunol.* **163**, 2937–2943 (1999).
- Denison, M. S., Fisher, J. M. & Whitlock, J. P. Jr. Inducible, receptor-dependent protein-DNA interactions at a dioxin-responsive transcriptional enhancer. *Proc. Natl Acad. Sci. USA* **85**, 2528–2532 (1988).
- Veldhoen, M., Hocking, R. J., Flavell, R. A. & Stockinger, B. Signals mediated by transforming growth factor- $\beta$  initiate autoimmune encephalomyelitis, but chronic inflammation is needed to sustain disease. *Nature Immunol.* **7**, 1151–1156 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Smith for providing us with the AHR-deficient mouse strain; A. Rae and G. Preece for cell sorting; and Biological Services at NIMR, especially T. Norton and H. Boyes, for animal care and EAE scoring. This work was funded by the Medical Research Council UK.

**Author Contributions** M.V. and K.H. performed the experiments; A.M.W. and J.B. did the microarrays and analysis; L.D. and J.-C.R. generated and provided the anti-IL-22 antibody; and B.S. directed the research and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.S. (bstock@nimr.mrc.ac.uk).



## METHODS

**Real-time PCR.** RNA was extracted using Trizol (Invitrogen) and 1-bromo-3-chloro-propane (Sigma) and reverse transcribed with oligo d(T)16 (Applied Biosystems) according to the manufacturer's protocol. The cDNA served as template for the amplification of genes of interest and the housekeeping gene (*Hprt*) by real-time PCR, using TaqMan Gene Expression Assays and Applied Biosystems 7900HT Fast Real-Time PCR System.

**Primers from Applied Biosystems.** Mouse: *Hprt*, Mm00446968\_m1; *Tbx21*, Mm00450960\_m1; *Gata3*, Mm00484683\_m1; *Foxp3*, Mm00475156\_m1; *Rorc*, Mm01261019\_g1; *Ahr*, Mm00478930\_m1; *Arnt*, Mm00507836\_m1; *Ahr* exon1 forward, 5'-CGCCTCCGGGACGCAGGTGG-3'; *Ahr* exon2 reverse, 5'-AAAG-AAGCTCTTGGCCCTCAG-3'; *Ifng*, Mm00801778\_m1; *Il4*, Mm00445259\_m1; *Il17a*, Mm00439619\_m1; *Il17f*, Mm00521423\_m1; *Il22*, Mm00444241\_m1; *Cyp1a1*, Mm00487217\_m1; *Il23r*, Mm00519942\_m1. Human: *HPRT*, HS99999909\_m1; *IL17A*, HS00174383\_m1; *IL17F*, HS00369400\_m1; *IL22*, HS00220924\_m1; *AHR*, HS00169233\_m1; *CYP1A1*, HS00153120\_m1.

**In vitro T-cell differentiation and cytokine staining.** Naive CD4<sup>+</sup> T cells were isolated by fluorescence-activated cell sorting (FACS) using a MoFlo sorter of lymph node cell suspensions for CD44<sup>lo</sup>CD25<sup>-</sup>CD4<sup>+</sup> cells. Natural T<sub>reg</sub> cells were sorted from lymph node suspensions as CD4<sup>+</sup>CD25<sup>+</sup> cells.

Conditions for different cell subsets were: T<sub>H</sub>0 (anti-IFN- $\gamma$  plus anti-IL-4 plus anti-TGF- $\beta$ ), T<sub>H</sub>1 (4 ng ml<sup>-1</sup> IL-12), T<sub>H</sub>2 (10 ng ml<sup>-1</sup> IL-4 and anti-IFN- $\gamma$ ), iT<sub>reg</sub> (10 ng ml<sup>-1</sup> TGF- $\beta$  and anti-IFN- $\gamma$ ) or T<sub>H</sub>17 (20 ng ml<sup>-1</sup> IL-6 plus 1 ng ml<sup>-1</sup> TGF- $\beta$  plus 10 ng ml<sup>-1</sup> IL-1 $\beta$  and anti-IFN- $\gamma$ ). Neutralizing antibodies were used at a concentration of 10  $\mu$ g ml<sup>-1</sup>: clone R46A2 anti-IFN- $\gamma$ , clone 11B11 anti-IL-4 and clone 1D11 anti-TGF- $\beta$ .

The AHR ligands FICZ (BioMol) or  $\beta$ -naphthoflavone (Sigma) were added in some experiments at the start of culture. For measurements of intracellular cytokines, T cells were re-stimulated with 500 ng ml<sup>-1</sup> phorbol dibutyrate and 500 ng ml<sup>-1</sup> ionomycin in the presence of brefeldin A for 4 h on day 5 after initiation of cultures. IL-17A antibody was obtained from eBioSciences; IL-17F antibody was from R&D. The anti-IL-22 antibody (MH22B2) was generated in IL-22 knockout BALB/c mice immunized with recombinant human IL-22 crosslinked to ovalbumin in the presence of glutaraldehyde and cross-reacts with human and murine IL-22.

**Retroviral transduction.** FACS-sorted naive CD4<sup>+</sup> T cells were plated in antibody-coated wells as described above in the absence of cytokines on day 0. On day 1, fresh retrovirus supernatant was added together with cytokines for polarization and the cells were spun for 1 h at 1,200 r.p.m. Thereafter, the cells were cultured in the presence or absence of 200 nM AHR ligand FICZ for another 4 days and then assayed by RT-PCR and intracellular staining.

**Generation of bone marrow chimaeras.** 5  $\times$  10<sup>6</sup> bone marrow cells from B6 donors or AHR-deficient donors were injected into lethally (9.5 Gy) irradiated AHR-deficient hosts (B6-AHR chimaeras) or B6 hosts (AHR- B6 chimaeras). Bone marrow donors were differentiated from the hosts by an allotypic marker (CD45.1). Chimaeras were tested for reconstitution 8 weeks after bone marrow transfer.

## LETTERS

# ***BCR-ABL1* lymphoblastic leukaemia is characterized by the deletion of Ikaros**

Charles G. Mullighan<sup>1</sup>, Christopher B. Miller<sup>1</sup>, Ina Radtke<sup>1</sup>, Letha A. Phillips<sup>1</sup>, James Dalton<sup>1</sup>, Jing Ma<sup>4</sup>, Deborah White<sup>5</sup>, Timothy P. Hughes<sup>5</sup>, Michelle M. Le Beau<sup>6</sup>, Ching-Hon Pui<sup>2</sup>, Mary V. Relling<sup>3</sup>, Sheila A. Shurtleff<sup>1</sup> & James R. Downing<sup>1</sup>

The Philadelphia chromosome, a chromosomal abnormality that encodes *BCR-ABL1*, is the defining lesion of chronic myelogenous leukaemia (CML) and a subset of acute lymphoblastic leukaemia (ALL)<sup>1–3</sup>. To define oncogenic lesions that cooperate with *BCR-ABL1* to induce ALL, we performed a genome-wide analysis of diagnostic leukaemia samples from 304 individuals with ALL, including 43 *BCR-ABL1* B-progenitor ALLs and 23 CML cases. *IKZF1* (encoding the transcription factor Ikaros) was deleted in 83.7% of *BCR-ABL1* ALL, but not in chronic-phase CML. Deletion of *IKZF1* was also identified as an acquired lesion at the time of transformation of CML to ALL (lymphoid blast crisis). The *IKZF1* deletions resulted in haploinsufficiency, expression of a dominant-negative Ikaros isoform, or the complete loss of Ikaros expression. Sequencing of *IKZF1* deletion breakpoints suggested that aberrant RAG-mediated recombination is responsible for the deletions. These findings suggest that genetic lesions resulting in the loss of Ikaros function are an important event in the development of *BCR-ABL1* ALL.

Acute lymphoblastic leukaemia (ALL) comprises a heterogeneous group of disorders characterized by recurring chromosomal abnormalities including translocations, trisomies and deletions. An ALL subtype with especially poor prognosis is characterized by the presence of the Philadelphia chromosome arising from the t(9;22)(q34;q11.2) translocation, which encodes the constitutively activated *BCR-ABL1* tyrosine kinase. *BCR-ABL1*-positive ALL constitutes 5% of paediatric B-progenitor ALL and approximately 40% of adult ALL<sup>1,2</sup>. Expression of *BCR-ABL1* is also the pathological lesion underlying CML<sup>3</sup>. Data from murine studies demonstrate that expression of *BCR-ABL1* in haematopoietic stem cells can alone induce a CML-like myeloproliferative disease, but cooperating oncogenic lesions are required for the generation of a blastic leukaemia<sup>4,5</sup>. Although the p210 and p190 *BCR-ABL1* fusions are most commonly found in CML and paediatric *BCR-ABL1* ALL, respectively, either fusion may be found in adult *BCR-ABL1* ALL<sup>6</sup>. Notably, a number of genetic lesions including additional cytogenetic aberrations and mutations in tumour suppressor genes have been described in CML cases progressing to blast crisis<sup>7</sup>. However, the specific lesions responsible for the generation of *BCR-ABL1* ALL and blastic transformation of CML remain incompletely understood<sup>7</sup>. To identify cooperating oncogenic lesions in ALL, we recently performed a genome-wide analysis of paediatric ALL<sup>8</sup>. This analysis identified an average of 6.8 genomic copy number alterations in 9 *BCR-ABL1* ALL cases, including deletions in genes that have a regulatory role in normal B-cell development.

To extend this analysis and identify lesions that distinguish CML from *BCR-ABL1* ALL, we have now examined DNA from leukaemic

samples from 304 paediatric and adult cases of ALL (254 B-progenitor; 50 T-lineage), including 21 paediatric and 22 adult *BCR-ABL1* ALL cases, and 23 adult CML cases (Supplementary Table 1). Samples were analysed using the 250K Sty and Nsp Affymetrix single-nucleotide polymorphism (SNP) arrays (and also the 100K arrays for most cases). This identified a mean of 8.79 somatic copy number alterations per *BCR-ABL1* ALL case (range 1–26), with 1.44 gains (range 0–13) and 7.33 losses (range 0–25) (Supplementary Table 4). No significant differences were noted in the frequency of copy number alterations between paediatric and adult *BCR-ABL1* ALL cases. The most frequent somatic copy number alteration was deletion of *IKZF1*, which encodes the transcription factor Ikaros (Table 1). *IKZF1* was deleted in 36 (83.7%) of 43 *BCR-ABL1* ALL cases, including 76.2% of paediatric and 90.9% of adult *BCR-ABL1* ALL cases. *CDKN2A* was deleted in 53.5% of *BCR-ABL1* ALL cases, most of which (87.5%) also had deletions of *IKZF1* (Table 1 and Supplementary Table 5). Conversely, of the *BCR-ABL1* ALL cases with *IKZF1* deletions, 41.6% lacked *CDKN2A* alterations. Deletion of *PAX5* occurred in 51% of *BCR-ABL1* ALL cases, again with the majority also having a deletion of *IKZF1* (95%) (Table 1 and Supplementary Table 5). No other defining copy number alterations were identified in the rare *BCR-ABL1* ALL cases that lacked a deletion of *IKZF1*.

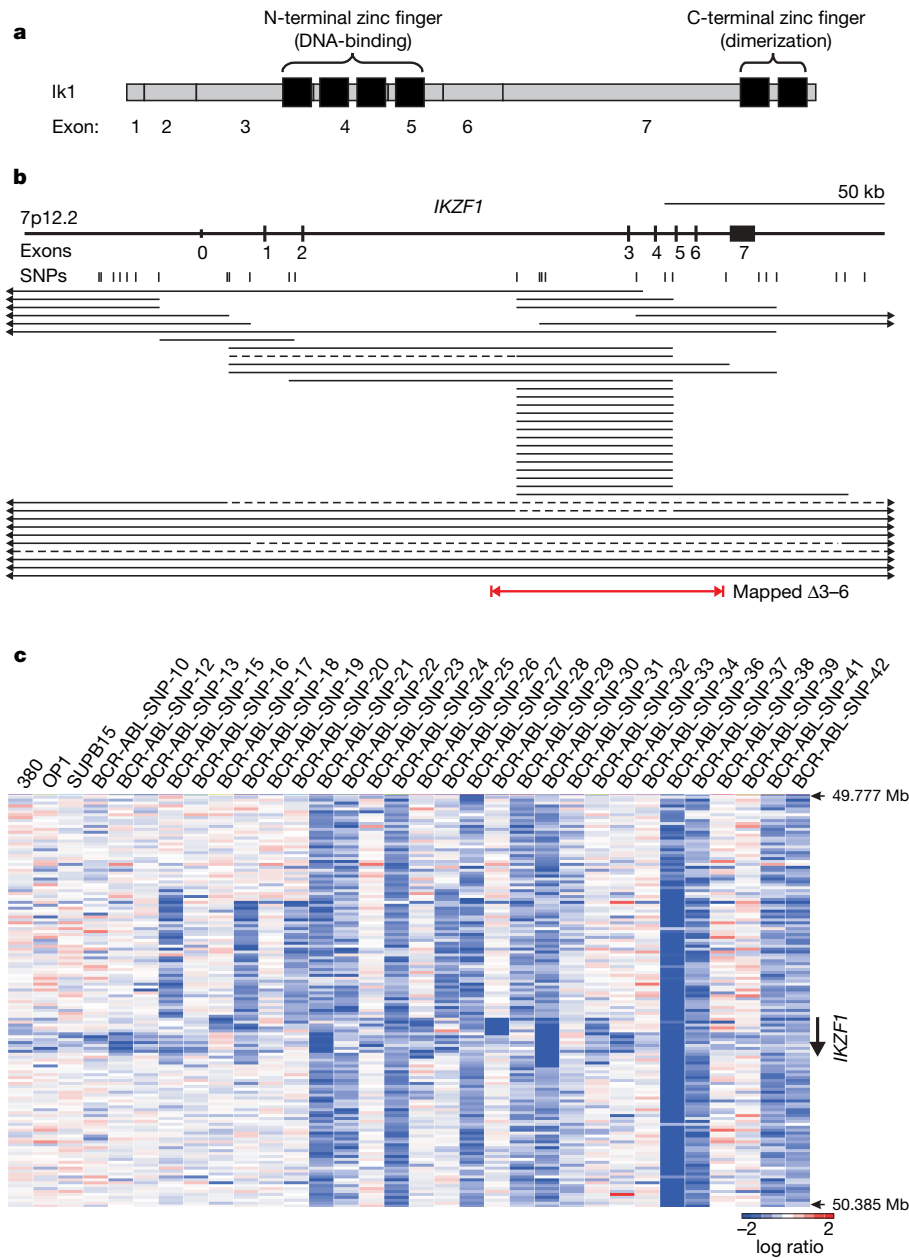
Ikaros is a member of a family of zinc-finger nuclear proteins that is required for normal lymphoid development<sup>9–12</sup>. Ikaros has a central DNA-binding domain consisting of four zinc fingers, and a homo- and heterodimerization domain consisting of the two carboxy-terminal zinc fingers<sup>13</sup> (Fig. 1 and Supplementary Fig. 1). Alternative splicing generates multiple Ikaros isoforms, several of which lack the amino-terminal zinc fingers required for DNA binding; however, the physiological relevance of these isoforms in normal haematopoiesis remains unclear<sup>9–11,14</sup> (Supplementary Fig. 1). The *IKZF1* deletions identified in *BCR-ABL1* ALL were predominantly mono-allelic and were limited to the gene in 25 cases, conclusively identifying *IKZF1* as the genetic target (Fig. 1). In 19 cases the deletions were confined to a subset of internal *IKZF1* exons, most commonly exons 3–6 ( $\Delta 3-6$ ;  $N = 15$ ). Notably, the  $\Delta 3-6$  deletion is predicted to encode an Ikaros isoform that lacks the DNA-binding domain but retains the C-terminal zinc fingers. The *IKZF1* deletions were confirmed by fluorescence *in situ* hybridization (FISH) and genomic quantitative polymerase chain reaction (PCR), and were in the predominant leukaemic clone (Supplementary Table 6 and Supplementary Fig. 2). Detailed analysis failed to reveal any evidence of either *IKZF1* point mutations or inactivation of its promoter by CpG methylation in primary ALL samples (data not shown and Supplementary Fig. 8).

<sup>1</sup>Departments of Pathology, <sup>2</sup>Oncology and <sup>3</sup>Pharmaceutical Sciences and <sup>4</sup>The Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>5</sup>Division of Haematology, The Institute for Medical and Veterinary Science, Adelaide, South Australia 5000, Australia. <sup>6</sup>Section of Hematology/Oncology, University of Chicago, Chicago, Illinois 60637, USA.

**Table 1 | Frequency of recurring DNA copy number abnormalities in ALL**

ALL subtype (N)	<i>IKZF1</i>	<i>CDKN2A</i>	<i>PAX5</i>	<i>C20orf94</i>	<i>RB1</i>	<i>MEF2C</i>	<i>EBF1</i>	<i>BTG1</i>	<i>DLEU</i>	<i>FHIT</i>	<i>ETV6</i>
B-progenitor (254)											
<i>BCR-ABL1</i> (43)	36	23	22	10	8	6	6	6	4	4	3
Childhood (21)	16	10	10	7	4	2	3	4	1	2	2
Adult (22)	20	13	12	3	4	4	3	2	3	2	1
Hypodiploid (10)	5	10	10	0	0	0	1	1	0	1	2
Other B ALLs (75)	15	25	22	4	1	0	2	5	1	3	10
High hyperdiploid (39)	2	8	4	1	3	0	0	0	5	0	3
<i>MLL</i> -rearranged (22)	1	4	4	0	2	0	0	0	3	0	2
<i>TCF3-PBX1</i> (17)	0	6	7	0	2	0	0	0	2	0	0
<i>ETV6-RUNX1</i> (48)	0	14	16	6	2	0	5	7	4	6	33
T-lineage (50)	2	36	5	1	6	1	3	0	3	0	4
Total (304)	61	126	90	22	24	7	17	19	22	14	57
<i>P</i> -value	$6.6 \times 10^{-27}$	$7.4 \times 10^{-10}$	$1.4 \times 10^{-9}$	$7.0 \times 10^{-8}$	$1.1 \times 10^{-6}$	0.0004	0.0247	$1.5 \times 10^{-7}$	$2.6 \times 10^{-6}$	0.0076	$9.1 \times 10^{-15}$

The prevalence of recurring genomic abnormalities in *BCR-ABL1* B-progenitor ALL identified by SNP array analysis is shown for each ALL subtype. The exact likelihood ratio *P*-value for variation in the frequency of each lesion across ALL subtypes is shown. The *DLEU* region at 13q14 incorporates the microRNA genes *MIRN16-1* and *MIRN15A*.



**Figure 1 | *IKZF1* deletions in *BCR-ABL1* ALL.** **a**, Domain structure of *IKZF1*. Exons 3–5 encode four N-terminal zinc fingers (black boxes) responsible for DNA binding. The C-terminal zinc fingers encoded by exon 7 are essential for homo- and heterodimerization. **b**, Genomic organization of *IKZF1* and location of each of the 36 deletions observed in *BCR-ABL1* B-progenitor ALL. Each line depicts the deletion(s) observed in each case. In four cases, two discontinuous

deletions were observed. Hemizygous deletions are solid lines and homozygous deletions dashed. Arrows indicate deletions extending beyond the limits of the figure. The exact boundaries of the deletions were defined by genomic quantitative PCR, and for *IKZF1* Δ3–6, by long-range genomic PCR (red arrow). **c**, dChip SNP raw log<sub>2</sub> ratio copy number data depicting *IKZF1* deletions for 29 *BCR-ABL1* cases and 3 B-progenitor ALL cell lines.

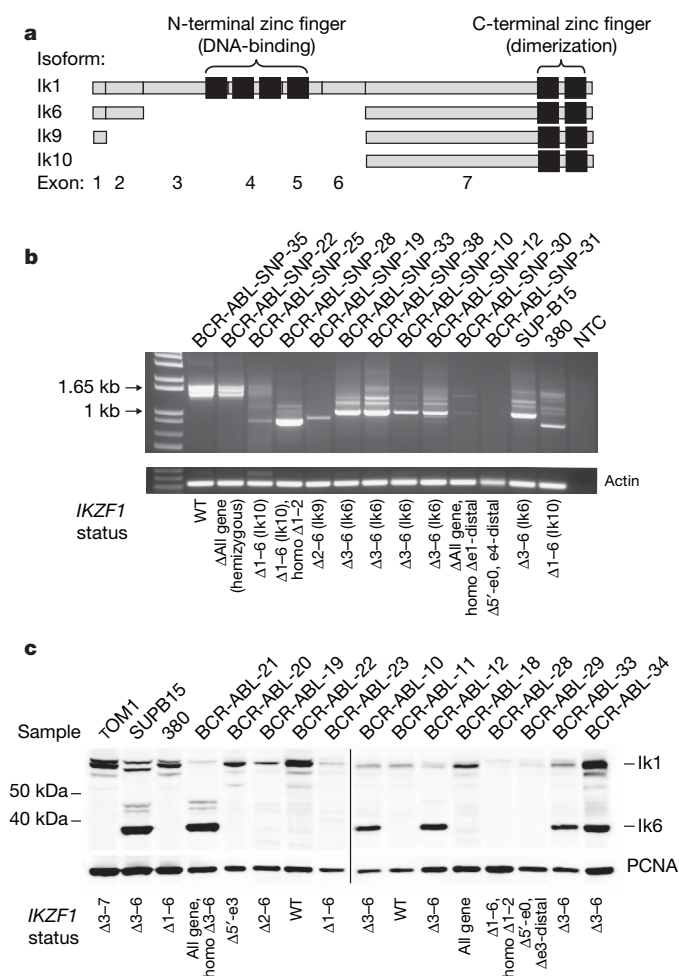


The expression of aberrant, dominant-negative Ikaros isoforms in B- and T-lineage ALL has been previously reported by several groups<sup>15–22</sup>, although alternative splicing has been reported to be the underlying mechanism<sup>23</sup>. Importantly, the  $\Delta 3-6$  isoform of Ikaros has been shown to function as a dominant-negative inhibitor of the transcriptional activity of Ikaros and related family members<sup>13</sup>. Moreover, mice homozygous for either an *Ikzf1* null mutation<sup>24</sup> or a dominant-negative *Ikzf1* mutation<sup>25</sup> exhibit profound defects in lymphoid development, and mice heterozygous for a dominant-negative *Ikzf1* mutation develop clonal T-cell expansions and lymphoproliferative diseases<sup>26</sup>, demonstrating that alteration in the level of *Ikzf1* expression is oncogenic.

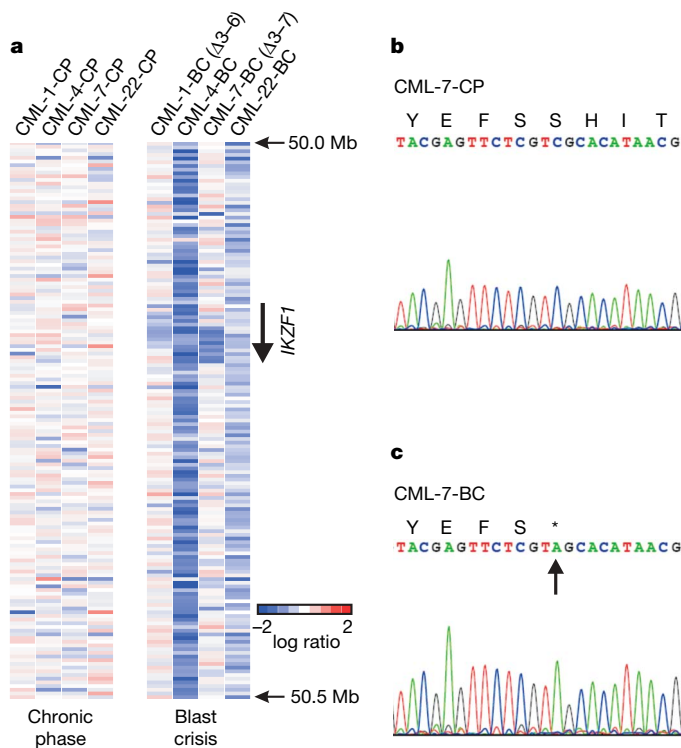
The high frequency of focal deletions in *IKZF1* in *BCR-ABL1* ALL suggests that expression of alternative *IKZF1* transcripts may be the result of specific genetic lesions, and not alternative splicing of an intact gene. To explore further this possibility, we performed reverse-transcriptase PCR (RT-PCR) analysis for *IKZF1* transcripts in 159

cases (Fig. 2). This demonstrated that expression of the Ik6 transcript, which lacks exons 3–6, was exclusively observed in cases harbouring the *IKZF1*  $\Delta 3-6$  deletion (Fig. 2b). Furthermore, we detected two previously unknown Ikaros isoforms exclusively in cases with larger deletions: Ik9 in a case with deletion of exons 2–6, and Ik10 in three cases with deletion of exons 1–6 (Fig. 2a, b, and Supplementary Fig. 3). For each isoform, Ik6, Ik9 and Ik10, there was concordance between the transcripts detected by RT-PCR and the extent of deletion defined by SNP array and genomic PCR analysis (Fig. 2b). Moreover, analysis of 22 *IKZF1*  $\Delta 3-6$  and 29 non- $\Delta 3-6$  cases with a quantitative RT-PCR assay specific for the Ik6 transcript confirmed that Ik6 expression was restricted to cases with the  $\Delta 3-6$  deletion ( $P = 6.41 \times 10^{-15}$ , Supplementary Fig. 4). Furthermore, the Ik6 protein isoform was only detectable by western blotting in cases with a  $\Delta 3-6$  *IKZF1* deletion (Fig. 2c). We also did not observe expression of Ik6 after the enforced expression of *BCR-ABL1* in *Arf* null or wild-type murine haematopoietic precursors (data not shown). Together, these data indicate that the expression of non-DNA-binding Ikaros isoforms is due to *IKZF1* genomic abnormalities, and not aberrant post-transcriptional splicing induced by *BCR-ABL1*, as has been suggested<sup>23</sup>.

To identify copy number alterations in CML, we performed SNP array analysis on 23 CML cases. In addition to chronic-phase CML (CP-CML), we also examined matched accelerated phase (AP-CML,  $N = 7$ ) and blast crisis (BC-CML,  $N = 15$  (12 myeloid and 3 lymphoid)) samples (Supplementary Table 2). This identified only 0.47 copy number alterations per CP-CML case (range 0–8) (Supplementary Table 7), suggesting that *BCR-ABL1* is sufficient to induce CML, but alone does not result in substantial genomic instability. Notably, no recurrent lesions were identified. In contrast, there was a mean of 7.8 copy number alterations per BC-CML case



**Figure 2 | Ikaros isoforms in ALL blasts.** **a**, Domain structure of the *IKZF1* isoforms detected by RT-PCR, examples of which are shown in **b**. **b**, RT-PCR for *IKZF1* transcripts (using exon 0- and 7-specific primers) in representative cases with various *IKZF1* genomic abnormalities. Each case expressing an aberrant isoform has a corresponding *IKZF1* genomic deletion. *IKZF1*  $\Delta 3-6$  was also detected in the *BCR-ABL1* ALL cell lines SUP-B15 and OP1, and  $\Delta 1-6$  in the ALL cell line 380. **c**, Western blotting for Ikaros using a C-terminus-specific polyclonal antibody. Ik6 was only detectable in cases with *IKZF1*  $\Delta 3-6$ . The  $\Delta 1-6$  and  $\Delta 2-6$  deletions do not produce a detectable protein. In three cases with multiple focal hemizygous deletions involving different regions of *IKZF1* (BCR-ABL-SNP-26, BCR-ABL-SNP-29 and BCR-ABL-SNP-31), no wild-type Ikaros was detectable by RT-PCR or western blotting, indicating that the deletions involve both copies of *IKZF1* in each case.



**Figure 3 | *IKZF1* deletions in blast crisis CML.** **a**, dChip SNP log<sub>2</sub> ratio copy number heatmaps of four CML cases showing acquisition of *IKZF1* deletions at progression to blast crisis. **b**, **c**, Pherograms of *IKZF1* exon 7 sequencing demonstrating acquisition of the coding nucleotide 1520C>A, amino acid Ser507X mutation at chronic-phase (**b**) and blast crisis (**c**) in case CML-7. As this case has a concomitant hemizygous *IKZF1* deletion involving exon 7, the mutation appears to be homozygous.

(range 0–28) (Supplementary Table 7), with *IKZF1* deletions in four BC samples, including two of the three cases with lymphoid blast crisis (Fig. 3a). Two of the *IKZF1* deletions involved the entire gene (CML-4-BC and CML-22-BC), one  $\Delta 3-6$  (CML-1-BC, which was associated with *Ik6* expression by RT-PCR) and one  $\Delta 3-7$  (CML-7-BC). CML-7-BC also had an *IKZF1* nonsense mutation in the C-terminal zinc-finger domain of exon 7 in the non-deleted allele (coding nucleotide 1520C>A, amino acid Ser507X, Fig. 3b, c). One BC sample had a *CDKN2A* deletion, and four cases had copy number alterations involving *PAX5* (two deletions, one internal amplification and one trisomy 9), with two of these also having *IKZF1* deletion. Copy number alterations were identified in two AP-CML samples. These data demonstrate an increased burden of genomic aberrations during progression of CML, with *IKZF1* mutation a frequent event in the transformation of CML to lymphoid blast crisis.

To explore the mechanism responsible for the identified *IKZF1* deletion, we sequenced the *IKZF1*  $\Delta 3-6$  genomic breakpoints (Supplementary Fig. 6). The deletions were restricted to highly localized sequences in introns 2 and 6 (Supplementary Fig. 7). Moreover, heptamer recombination signal sequences (RSSs) recognized by the RAG enzymes during V(D)J recombination<sup>27</sup> were located immediately internal to the deletion breakpoints, and a variable number of additional nucleotides were present between the consensus intron 2 and 6 sequences, suggestive of the action of terminal deoxynucleotidyl transferase (TdT). Together, these data suggest that the *IKZF1*  $\Delta 3-6$  deletion arises owing to aberrant RAG-mediated recombination.

We have identified a high frequency of copy number alterations in *BCR-ABL1* ALL and BC-CML, but not in CP-CML. We observed a near obligate deletion of *IKZF1* in *BCR-ABL1* ALL, with 83.7% of paediatric and adult cases containing deletions that lead to a reduction in dose and/or the expression of an altered Ikaros isoform. By contrast, deletion of *IKZF1* was not detected in CP-CML, but was identified as an acquired lesion in two of three lymphoid BC-CML samples. These data, together with the low frequency of *IKZF1* deletions in other paediatric B-progenitor ALL cases, and the lack of focal *IKZF1* aberrations in recently reported genomic analysis of non-haematopoietic tumours<sup>28</sup> (Supplementary Results), suggest that alterations in Ikaros directly contribute to the pathogenesis of *BCR-ABL1* ALL. How reduced activity of Ikaros, and possibly that of other family members through the expression of dominant negative Ikaros isoforms, collaborates with *BCR-ABL1* to induce lymphoblastic leukaemia remains to be determined. Importantly, mice with attenuated Ikaros expression exhibit a partial block of B lymphoid maturation at the pro-B-cell stage<sup>29</sup>, suggesting that Ikaros loss may contribute to the arrested B-lymphoid maturation in *BCR-ABL1* ALL. However, the high co-occurrence of *PAX5* deletions in many cases suggests that *IKZF1* deletion contributes to transformation in additional ways. The frequent co-deletion of *CDKN2A* (encoding *INK4A/ARF*) with *IKZF1* in *BCR-ABL1* ALL is a notable finding. This suggests that attenuated Ikaros activity may either collaborate with disruption of *INK4A/ARF*-mediated tumour suppression, or act through alternative uncharacterized tumour suppressor pathways in ALL. Furthermore, the identification of aberrant RAG-mediated recombination as the mechanism underlying deletions of *IKZF1* suggests that the cellular target of this transforming event is downstream of the haematopoietic stem cell. Dissecting the contribution of altered Ikaros activity to *BCR-ABL1* leukaemogenesis should not only provide valuable mechanistic insights, but will also help to determine if the presence of this genetic lesion can be used to gain a therapeutic advantage against this aggressive leukaemia.

## METHODS SUMMARY

Two-hundred and eighty-two paediatric ALL cases, 22 adult *BCR-ABL1* ALL cases, 49 samples obtained from 23 adult patients with chronic myeloid leukaemia (CML) and 36 leukaemia cell lines were studied (Supplementary Tables 1 and 2). Affymetrix 250K Sty and Nsp arrays were performed on all samples. 50K Hind 240 and 50K Xba 240 arrays were performed for 252 ALL samples

(Supplementary Table 1). SNP array data were analysed using dChip (<http://www.dChip.org>), a reference normalization algorithm, and circular binary segmentation as previously described<sup>8</sup>. *IKZF1* deletions were confirmed by FISH and/or genomic quantitative PCR. Expression of Ikaros transcripts was examined by qualitative and quantitative RT-PCR, and western blotting. Genomic sequencing was performed for all *IKZF1* coding exons. Methylation status of the *IKZF1* promoter CpG island was performed by MALDI-TOF mass spectrometry of bisulphite-treated leukaemic blast DNA.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 18 December 2007; accepted 25 February 2008.

Published online 13 April 2008.

- Ribeiro, R. C. *et al.* Clinical and biologic hallmarks of the Philadelphia chromosome in childhood acute lymphoblastic leukemia. *Blood* **70**, 948–953 (1987).
- Gleissner, B. *et al.* Leading prognostic relevance of the BCR-ABL translocation in adult acute B-lineage lymphoblastic leukemia: a prospective study of the German Multicenter Trial Group and confirmed polymerase chain reaction analysis. *Blood* **99**, 1536–1543 (2002).
- Goldman, J. M. & Melo, J. V. Chronic myeloid leukemia—advances in biology and new approaches to treatment. *N. Engl. J. Med.* **349**, 1451–1464 (2003).
- Daley, G. Q., Van Etten, R. A. & Baltimore, D. Blast crisis in a murine model of chronic myelogenous leukemia. *Proc. Natl Acad. Sci. USA* **88**, 11335–11338 (1991).
- Williams, R. T., Roussel, M. F. & Sherr, C. J. Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **103**, 6688–6693 (2006).
- Melo, J. V. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* **88**, 2375–2384 (1996).
- Melo, J. V. & Barnes, D. J. Chronic myeloid leukaemia as a model of disease evolution in human cancer. *Nature Rev. Cancer* **7**, 441–453 (2007).
- Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
- Hahn, K. *et al.* The lymphoid transcription factor Lyf-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. *Mol. Cell. Biol.* **14**, 7111–7123 (1994).
- Molnar, A. & Georgopoulos, K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol. Cell. Biol.* **14**, 8292–8303 (1994).
- Molnar, A. *et al.* The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. *J. Immunol.* **156**, 585–592 (1996).
- Rebollo, A. & Schmitt, C. Ikaros, Aiolo and Helios: transcription regulators and lymphoid malignancies. *Immunol. Cell Biol.* **81**, 171–175 (2003).
- Sun, L., Liu, A. & Georgopoulos, K. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* **15**, 5358–5369 (1996).
- Klug, C. A. *et al.* Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. *Proc. Natl Acad. Sci. USA* **95**, 657–662 (1998).
- Sun, L. *et al.* Expression of dominant-negative Ikaros isoforms in T-cell acute lymphoblastic leukemia. *Clin. Cancer Res.* **5**, 2112–2120 (1999).
- Sun, L. *et al.* Expression of aberrantly spliced oncogenic Ikaros isoforms in childhood acute lymphoblastic leukemia. *J. Clin. Oncol.* **17**, 3753–3766 (1999).
- Sun, L. *et al.* Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **96**, 680–685 (1999).
- Nakase, K. *et al.* Dominant negative isoform of the Ikaros gene in patients with adult B-cell acute lymphoblastic leukemia. *Cancer Res.* **60**, 4062–4065 (2000).
- Olivero, S. *et al.* Detection of different Ikaros isoforms in human leukaemias using real-time quantitative polymerase chain reaction. *Br. J. Haematol.* **110**, 826–830 (2000).
- Nishii, K. *et al.* Expression of B cell-associated transcription factors in B-cell precursor acute lymphoblastic leukemia cells: association with PU.1 expression, phenotype, and immunogenotype. *Int. J. Hematol.* **71**, 372–378 (2000).
- Takanashi, M. *et al.* Expression of the Ikaros gene family in childhood acute lymphoblastic leukaemia. *Br. J. Haematol.* **117**, 525–530 (2002).
- Tonnelle, C. *et al.* Overexpression of dominant-negative Ikaros 6 protein is restricted to a subset of B common adult acute lymphoblastic leukemias that express high levels of the CD34 antigen. *Hematol. J.* **4**, 104–109 (2003).
- Klein, F. *et al.* BCR-ABL1 induces aberrant splicing of IKAROS and lineage infidelity in pre-B lymphoblastic leukemia cells. *Oncogene* **25**, 1118–1124 (2006).
- Wang, J. H. *et al.* Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. *Immunity* **5**, 537–549 (1996).
- Georgopoulos, K. *et al.* The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143–156 (1994).
- Winandy, S., Wu, P. & Georgopoulos, K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell* **83**, 289–299 (1995).
- Fugmann, S. D. *et al.* The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).

28. Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898 (2007).
29. Kirstetter, P. *et al.* Ikaros is critical for B cell differentiation and function. *Eur. J. Immunol.* **32**, 720–730 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The authors thank Z. Cai for technical help, K. Rakestraw and J. Armstrong for assistance with sequencing, R. Williams and C. Sherr for the provision of Arf null hematopoietic cells and *BCR-ABL1* retroviral vectors, O. Heidenreich for providing the SKNO-1 cell line, and D. Campana for providing the OP1 cell line. This study was supported by the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital. C.G.M. was supported by grants from the National Health and Medical Research Council

(Australia), the Royal Australasian College of Physicians, and the Haematology Society of Australasia.

**Author Contributions** C.G.M. collected and extracted clinical samples, performed laboratory assays and analysed data. C.B.M., L.A.P., J.D. and I.R. performed laboratory assays. J.M. analysed SNP array data. D.W., T.P.H., M.M.L., C.-H.P., M.V.R. and S.A.S. collected clinical samples and data. C.G.M. and J.R.D. designed the study and wrote the manuscript, which was reviewed by all authors.

**Author Information** The primary SNP microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession numbers GSE9109–GSE9113. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.R.D. ([james.downing@stjude.org](mailto:james.downing@stjude.org)).



## METHODS

**Patients and samples.** Patients and samples comprised 282 patients with acute lymphoblastic leukaemia (ALL) treated at St Jude Children's Research Hospital, 22 adult *BCR-ABL1* ALL patients treated at the University of Chicago, and 49 samples obtained from 23 adult patients with chronic myeloid leukaemia (CML) treated at the Institute of Medical and Veterinary Science, Adelaide (Supplementary Tables 1 and 2). The CML cohort included 24 chronic phase, 7 accelerated phase and 15 blast crisis samples, and three samples obtained at complete cytogenetic response. All blast crisis samples were flow sorted to at least 90% blast purity before DNA extraction using FACSVantage s.e. (with DiVa option) flow cytometers (BD Biosciences) and fluorescein-isothiocyanate-labelled CD45, allophycocyanin-labelled CD33 and phycoerythrin-labelled CD19 and CD13 antibodies (BD Biosciences). Germline tissue was obtained by also sorting the non-blast population in seven cases. Informed consent for the use of leukaemic cells for research was obtained from patients, parents or guardians in accordance with the Declaration of Helsinki, and study approval was obtained from the SJCRH institutional review board.

**Cell lines examined by SNP array.** Thirty-six acute myeloid and lymphoid leukaemia cell lines were genotyped using the Affymetrix Mapping 250k Sty and Nsp arrays. These were the ALL cell lines 380 (*MYC-IGH* and *BCL2-IGH* B precursor), 697 (*TCF3-PBX1*), AT1 (*ETV6-RUNX1*), BV173 (CML in lymphoid blast crisis), CCRF-CEM (*TAL-SIL*), Jurkat (T-ALL), Kasumi-2 (*TCF3-PBX1*), MHH-CALL-2 (hyperdiploid B-precursor ALL), MHH-CALL-3 (*TCF3-PBX1*), MOLT3 (T-ALL), MOLT4 (T-ALL), NALM-6 (B-precursor ALL), OP1 (*BCR-ABL1*), Reh (*ETV6-RUNX1*), RS4;11 (*MLL-AF4*), SD1 (*BCR-ABL1*), SUP-B15 (*BCR-ABL1*), TOM-1 (*BCR-ABL1*), U-937 (*PICALM-AF10*), UOCB1 (*TCF3-HLF*), YT (NK leukaemia), and the AML cell lines CMK (FAB M7), HL-60 (FAB M2), K-562 (CML in myeloid blast crisis), Kasumi-1 (*RUNX1-RUNX1T1*), KG-1 (myelocytic leukaemia), ME-1 (*CBFB-MYH11*), ML-2 (*MLL-AF6*), M-07e (FAB M7), Mono Mac 6 (*MLL-AF9*), MV4-11 (*MLL-AF4*), NB4 (*PML-RARA*), NOMO-1 (*MLL-AF9*), PL21 (FAB M3), SKNO-1 (*RUNX1-RUNX1T1*) and THP-1 (FAB M5). Cell lines were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany, the American Type Culture Collection, Manassas, Virginia, from local institutional repositories, or were gifts from O. Heidenreich (SKNO-1) and D. Campana (OP1). Cells were cultured in accordance with previously published recommendations<sup>30</sup>. The paediatric *BCR-ABL1* B-precursor ALL cell line OP1 (ref. 31) was cultured in RPMI-1640 containing 100 units ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin, 2 mM glutamine and 10% fetal bovine serum. DNA was extracted from 5 × 10<sup>6</sup> cells obtained during log-phase growth after washing in PBS using the QIamp DNA blood mini kit (Qiagen).

**SNP microarray analysis.** Collection and processing of diagnostic and remission bone marrow and peripheral blood samples for Affymetrix SNP microarray analysis has been previously reported in detail<sup>8</sup>. Affymetrix 250K Sty and Nsp arrays were performed on all samples. 50K Hind 240 and 50K Xba 240 arrays were performed for 252 ALL samples (Supplementary Table 1). SNP array .CEL and SNP call .TXT files (generated by Affymetrix GTYPE 4.0 using the DM

algorithm) have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO series accession numbers GSE9109–GSE9113. These accessions contain the following data: GSE9109, Sty and Nsp files for 304 ALL samples, and Hind and Xba files for 252 of these samples; GSE9110, Sty and Nsp files for 56 CML samples; GSE9111, Sty, Nsp, Hind and Xba files for 50 remission acute leukaemia samples used as references for copy number analysis; GSE9112, Sty and Nsp files for 36 acute leukaemia cell lines; GSE9113, a superseries containing all of the above data. The data are also available at <http://www.stjude-research.org/data/ALL-SNP2/>.

**FISH.** FISH for *IKZF1* deletion was performed using diagnostic bone marrow or peripheral blood leukaemic cells in Carnoy's fixative as previously described<sup>8</sup>. BAC clones CTD-2382L6 and CTC-791O3 (for *IKZF1*, Open Biosystems) were labelled with fluorescein isothiocyanate, and control 7q31 probes RP11-460K21 (Children's Hospital Oakland Research Institute) and CTB-133K23 (Open Biosystems) were labelled with rhodamine. At least 100 interphase nuclei were scored per case.

**IKZF1 PCR, cloning, quantitative PCR and genomic sequencing.** RNA was extracted and reverse transcribed using random hexamer primers and Superscript III (Invitrogen) as previously described<sup>8</sup>. *IKZF1* transcripts were amplified from cDNA using the Advantage 2 PCR enzyme (Clontech) as previously described<sup>8</sup> using primers that anneal in exon 0 and 7 of *IKZF1*. PCR products were purified and sequenced directly and after cloning into pGEM-T-Easy (Promega). Genomic quantitative PCR for exons 1–7 of *IKZF1*, and real-time PCR to quantify expression of *Ik6*, were performed as previously described<sup>8</sup>. All primers and probes are listed in Supplementary Table 3. Genomic sequencing of *IKZF1* exons 0–7 in all ALL and CML samples was performed as previously described<sup>8</sup>.

**Western blotting.** Whole-cell lysates of 3–6 × 10<sup>6</sup> leukaemic cells were prepared and blotted as previously described<sup>8</sup> using N- and C-terminus-specific rabbit polyclonal Ikaros antibodies (Santa Cruz Biotechnology).

**Methylation analysis.** Methylation status of the *IKZF1* promoter CpG island (PCR amplicon hg17 coordinates: chromosome 7 50121508–50121714) was performed using MALDI-TOF mass spectrometry of PCR-amplified, bisulphite-modified genomic DNA extracted from leukaemic cells as previously described<sup>8,32</sup>.

**Statistical analysis.** Associations between ALL subtype and *IKZF1* deletion frequency were calculated using the exact likelihood ratio test. Differences in *Ik6* expression between *IKZF1* Δ3–6 and non-Δ3–6 cases were assessed using the exact Wilcoxon–Mann–Whitney test. All *P*-values reported are two-sided. Analyses were performed using StatXact v8.0.0 (Cytel).

30. Drexler, H. G. *The Leukemia-Lymphoma Cell Line Facts Book* 1st edn (Academic Press, London, 2001).
31. Manabe, A. *et al.* Interleukin-4 induces programmed cell death (apoptosis) in cases of high-risk acute lymphoblastic leukemia. *Blood* **83**, 1731–1737 (1994).
32. Ehrlich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl Acad. Sci. USA* **102**, 15785–15790 (2005).

# Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions

Stavroula Mili<sup>1</sup>, Konstadinos Moissoglu<sup>2</sup> & Ian G. Macara<sup>1</sup>

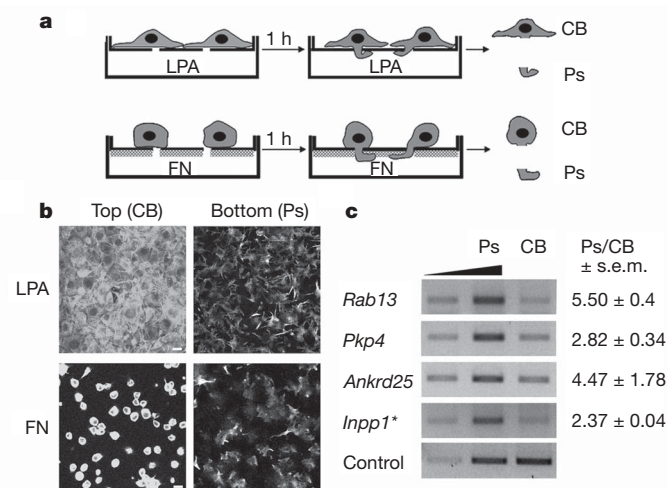
RNA localization is important for the establishment and maintenance of polarity in multiple cell types. Localized RNAs are usually transported along microtubules or actin filaments<sup>1</sup> and become anchored at their destination to some underlying subcellular structure. Retention commonly involves actin or actin-associated proteins<sup>2–7</sup>, although cyokeratin filaments and dynein anchor certain RNAs<sup>8,9</sup>. RNA localization is important for diverse processes ranging from cell fate determination to synaptic plasticity; however, so far there have been few comprehensive studies of localized RNAs in mammalian cells. Here we have addressed this issue, focusing on migrating fibroblasts that polarize to form a leading edge and a tail in a process that involves asymmetric distribution of RNAs<sup>10–12</sup>. We used a fractionation scheme<sup>13</sup> combined with microarrays to identify, on a genome-wide scale, RNAs that localize in protruding pseudopodia of mouse fibroblasts in response to migratory stimuli. We find that a diverse group of RNAs accumulates in such pseudopodial protrusions. Through their 3' untranslated regions these transcripts are anchored in granules concentrated at the plus ends of deetyrosinated microtubules. RNAs in the granules associate with the adenomatous polyposis coli (APC) tumour suppressor and the fragile X mental retardation protein (FMRP). APC is required for the accumulation of transcripts in protrusions. Our results suggest a new type of RNA anchoring mechanism as well as a new, unanticipated function for APC in localizing RNAs.

To identify on a genome-wide scale RNAs that are enriched at the leading edge of migrating cells, we used a fractionation method in which cells are plated on a microporous filter and then induced to polarize and extend pseudopodial protrusions in response to a migratory stimulus. Pseudopodia and cell bodies are then physically isolated and their contents compared<sup>13</sup>. We isolated pseudopodia and cell body fractions from NIH/3T3 cells extending protrusions in response to a chemotactic (lysophosphatidic acid, LPA) or a haptotactic (fibronectin) stimulus (Fig. 1a, b and Supplementary Fig. 1a). The quality of the fractionation was assessed by immunoblotting for activated focal adhesion kinase (FAK, phosphorylated at Y397), which is concentrated at the leading edge during migration<sup>13,14</sup>. Activated FAK was enriched in the pseudopodial fraction (Supplementary Fig. 1b), whereas total FAK and Ran were not. Total RNA from pseudopodia and cell body fractions was hybridized on Affymetrix GeneChip arrays, which analyse the expression level of over 39,000 transcripts. The resulting signals were processed to identify transcripts that show an asymmetric distribution.

About 50 RNAs were significantly enriched in pseudopodia in response to both migratory stimuli (Supplementary Table 1). This enrichment was verified by quantitative polymerase chain reaction with reverse transcription (RT–PCR; Fig. 1c) and real-time RT–PCR (Supplementary Fig. 1c) analyses of representative RNAs. Notably, these RNAs did not include  $\beta$ -actin or *Arp2/3* subunit transcripts,

which have been previously observed at lamellipodial regions<sup>10,12</sup>, possibly because in some cell types these RNAs accumulate at the leading edge in only a small percentage of cells<sup>15</sup>. Most transcripts enriched in pseudopodia encode proteins with various functions, ranging from membrane traffic to cytoskeletal organization, signalling, microtubule-based transport and RNA metabolism, as well as a number of uncharacterized proteins (Supplementary Table 1). However, comparison of the primary structure of these transcripts did not reveal any readily identifiable motifs shared among them, suggesting that potential localization signals are probably defined by a combination of primary structure and higher order structure, as is common for most localized RNAs<sup>16</sup>.

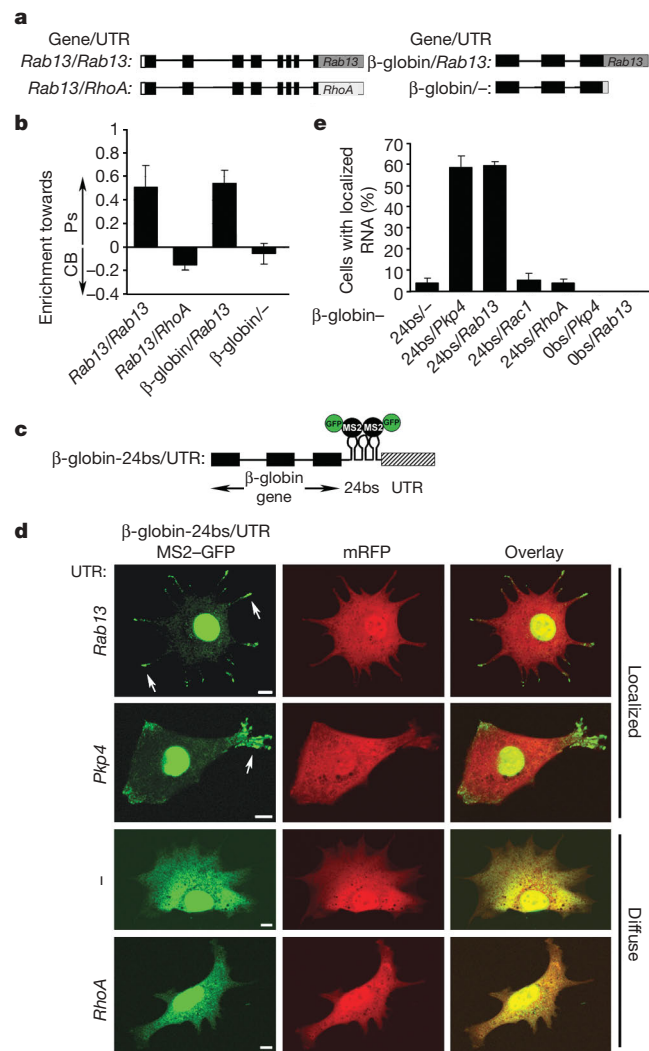
To dissect the localization mechanisms, we focused on the *Rab13* and plakophilin 4 (*Pkp4*) messenger RNAs, which both showed



**Figure 1 | Several RNAs are enriched in protruding pseudopodia of migrating cells.** **a**, Schematic diagram depicting strategies used for isolation of pseudopodia (Ps) and cell bodies (CB). NIH/3T3 cells, plated on microporous filters, were induced to extend protrusions by adding LPA in the bottom chamber. Alternatively, the underside of microporous filters was coated with fibronectin (FN) and cells plated on top extended protrusions towards the fibronectin-coated surface. Cell body and pseudopodia fractions were subsequently isolated. **b**, Cells extending protrusions in response to LPA or fibronectin, as described in **a**, were stained with fluorescein-isothiocyanate-conjugated phalloidin. Confocal images of the top and bottom side of the filter are shown. Scale bar, 15  $\mu$ m. **c**, Total RNA from pseudopodia and cell body fractions of cells extending protrusions in response to LPA was analysed by RT–PCR to detect the mRNAs indicated on the left. Increasing amounts of the pseudopodia sample were amplified (lanes 1 and 2) to ensure linearity of the amplification. Values on the right indicate mean pseudopodia/cell body ratios, normalized to the control *Arpc3* mRNA,  $\pm$  s.e.m.,  $n = 3$ .

<sup>1</sup>Department of Microbiology, Center for Cell Signaling, <sup>2</sup>Cardiovascular Research Center, University of Virginia, HSC, Charlottesville, Virginia 22908-0577, USA.

robust localization. First, we expressed in NIH/3T3 cells the *Rab13* gene encompassing the whole open reading frame and 3' untranslated region (UTR) (Fig. 2a). A Flag tag at the 5' end was used to distinguish the exogenous mRNA from endogenous transcript. Transfected cells were induced to extend pseudopodial protrusions in response to LPA, and pseudopodia and cell body fractions were isolated. RT-PCR analysis revealed that, like the endogenous *Rab13* mRNA, the reporter transcript was enriched in pseudopodia (Fig. 2b). Pseudopodial enrichment was abolished by replacement of the 3' UTR of the *Rab13* gene with the corresponding region from the non-localized *RhoA* gene. Moreover, the non-localized  $\beta$ -globin mRNA was recruited to pseudopodia when attached to the *Rab13* 3' UTR (Fig. 2a, b). All exogenous RNAs were expressed at similar



**Figure 2 | The 3' UTRs direct RNAs to accumulate in granules at tips of protrusions.** **a**, Schematic of transfected constructs. Black boxes, exons; black lines, introns; grey boxes, 3' UTR; white box, Flag tag. **b**, Cells transfected with the constructs depicted in **a** were fractionated into pseudopodia and cell body fractions after induction with LPA, and RNA enrichment towards each fraction was calculated ( $n = 2-4$ ). Error bars, s.e.m. **c**, Schematic depicting the general structure of  $\beta$ -globin constructs used in **d** and **e**. 24bs, 24 MS2-binding sites. **d**, Imaging of live cells co-transfected with plasmids encoding mRFP, MS2-GFP and  $\beta$ -globin constructs with various 3' UTRs. Shown are representative examples of the localization patterns observed when  $\beta$ -globin mRNA carried the UTRs indicated on the left (localized in granules at tips of protrusions (arrows) or diffuse in the cytoplasm). Scale bar, 10  $\mu$ m. **e**, Quantification of the percentage of cells exhibiting localized RNA distribution when transfected as in **d**, with constructs carrying the indicated 3' UTRs ( $n = 3$ ). Error bars, s.d.

levels, close to the expression level of the endogenous *Rab13* mRNA (Supplementary Fig. 2a, b), and in all experiments the distribution of the endogenous *Rab13* mRNA was determined in parallel, to ensure the reproducibility of the fractionation (data not shown). We conclude that the *Rab13* 3' UTR is necessary and sufficient to direct RNA accumulation in pseudopodia.

To visualize the localization pattern conferred by the *Rab13* 3' UTR, we used the MS2 system<sup>17</sup>. The MS2 bacteriophage coat protein binds with high affinity to RNA elements, multiple repeats of which are introduced into a reporter transcript. MS2 is fused to green fluorescent protein (GFP) and carries a nuclear localization signal to force accumulation in the nucleus. When co-expressed, the MS2-GFP binds to the reporter RNA and accompanies it to the cytoplasm, thus providing indirect detection of reporter RNA distribution in the cytoplasm of live cells.

Twenty-four repeats of the MS2-binding site were introduced into the  $\beta$ -globin gene downstream of the coding region, followed by different 3' UTR sequences (Fig. 2c). These constructs were co-expressed with MS2-GFP and fluorescence was monitored in live cells during spreading on a fibronectin-coated surface. Cells expressing  $\beta$ -globin mRNA with control 3' UTRs derived from the vector or from two non-localized mRNAs, *Rac1* and *RhoA*, exhibited diffuse fluorescence throughout the cytoplasm. The GFP signal overlapped entirely with the fluorescence from co-transfected mRFP protein, used as a diffuse cytosolic marker (Fig. 2d, bottom panels, and e, constructs  $\beta$ -globin-24bs/-,  $\beta$ -globin-24bs/*Rac1* and  $\beta$ -globin-24bs/*RhoA*). In contrast, in a large proportion of the cells that expressed a  $\beta$ -globin mRNA carrying the *Rab13* 3' UTR, GFP was concentrated in granules at the tips of protrusions (Fig. 2d, arrows, top panels, and e, construct  $\beta$ -globin-24bs/*Rab13*). Granular accumulation did not result from differences in expression levels, as all RNAs were expressed at similar amounts (Supplementary Fig. 2c). Furthermore, the signal was dependent on the presence of the 24 MS2-binding sites in the mRNA (Fig. 2e, construct  $\beta$ -globin-0bs/*Rab13*), confirming that it reflects the distribution of the transfected RNA. Significantly, a similar localization was conferred by the 3' UTR of *Pkp4* (Fig. 2d, e, construct  $\beta$ -globin-24bs/*Pkp4*). Therefore, certain 3' UTRs can direct RNAs to granules at the tips of protrusions, and this property is shared among mRNAs enriched in pseudopodia.

When imaged over time, the localized RNA granules are relatively stationary (Supplementary Movie 1) and appear and disappear with surprisingly slow kinetics (Supplementary Fig. 3). This stable accumulation of RNAs could represent either the steady state of a dynamic movement of RNA molecules to and from the granules, or result from sequestration into some cellular structure. To distinguish between these possibilities, we performed fluorescence recovery after photobleaching (FRAP). Because the RNA granules vary in size (see Fig. 2d and Supplementary Fig. 3), we focused, for these experiments, on the smaller granules of approximately 1–2  $\mu$ m in diameter. Whereas bleaching of the fluorescence signal in internal cytoplasmic areas was followed by rapid recovery (Fig. 3a), bleaching of the localized RNA granules was followed by very slow fluorescence recovery (Fig. 3a and Supplementary Movies 2 and 3), indicating that the transcripts present in these granules are stably associated/anchored in these structures and do not exchange rapidly with free cytoplasmic RNA molecules.

To gain insight into the identity of the structures that anchor transcripts at the tips of protrusions, the localized RNAs were expressed together with fluorescently tagged markers of various cellular structures. The RNA granules did not co-localize significantly with DCP1-containing P-bodies, focal adhesions (Supplementary Fig. 4a, c, d), cortical actin or actin stress fibres (Supplementary Fig. 4b). Notably, the granules appeared to be associated with microtubules, visualized through expression of RFP-tubulin, and were specifically concentrated at their plus ends (Fig. 3b and Supplementary Fig. 5). Consistent with a microtubule association, the

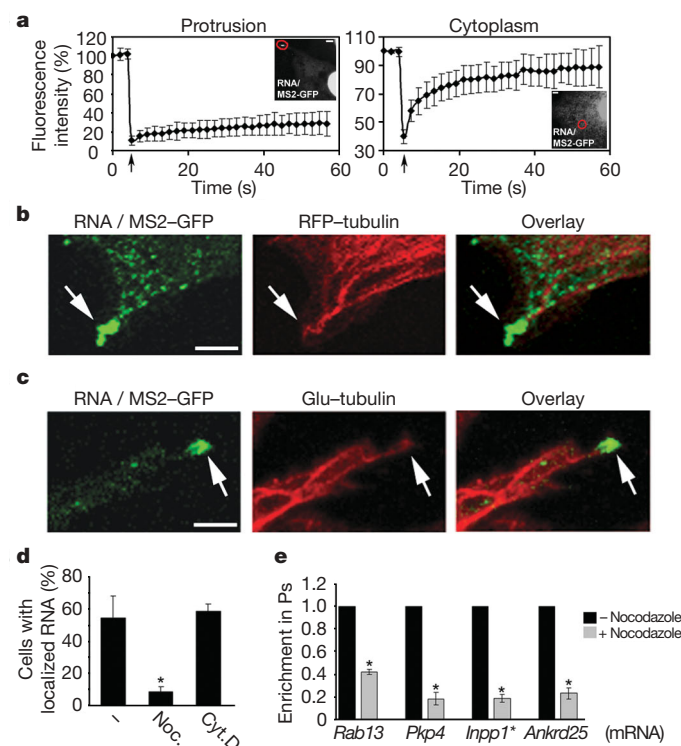


RNA granules largely disappeared when cells were treated with the microtubule-depolymerizing agent nocodazole (Fig. 3d and Supplementary Fig. 6b), at concentrations that do not affect the actin cytoskeleton (Supplementary Fig. 6a). In contrast, the presence of RNA granules was not affected when the actin cytoskeleton was disrupted by cytochalasin D (Fig. 3d and Supplementary Fig. 6a, b). Notably, brief treatment with nocodazole significantly reduced the pseudopodial enrichment of four different endogenous localized RNAs (Fig. 3e), without affecting the overall integrity and number of protrusions (Supplementary Fig. 6c). We conclude that association with microtubule plus ends is a property shared by multiple RNAs found in pseudopodia.

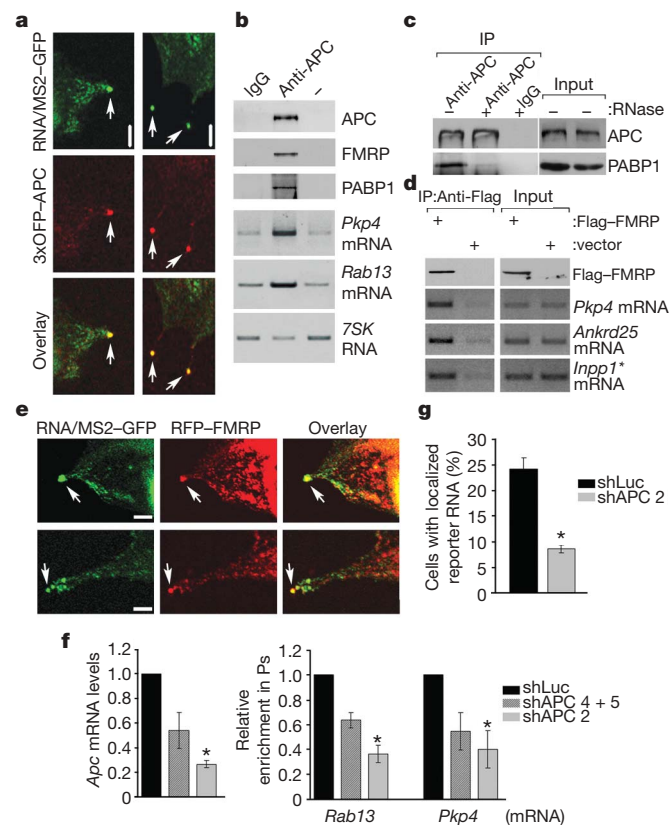
The fact that the RNA granules remain stationary over several minutes (Supplementary Movie 1) suggested that the microtubules with which they associate do not exhibit dynamic instability. Indeed, the RNA granules did not co-localize significantly with EB1-RFP or RFP-CLIP170 (Supplementary Fig. 7a, b), two plus-end tracking proteins known to associate with plus ends of dynamic microtubules<sup>18,19</sup>. Furthermore, FRAP analysis on protrusions of EB1-GFP-expressing cells showed that, in contrast to the RNA granules, EB1 is highly dynamic with fluorescence recovering after

a few seconds (Supplementary Fig. 8a and Supplementary Movie 4). Therefore, we tested whether the RNA granules associate specifically with the plus ends of stable microtubules, which are marked by post-translational modifications of tubulin. Although the RNA granules were not attached to acetylated microtubules (Supplementary Fig. 9a), they did associate with the plus ends of detyrosinated microtubules (Glu-microtubules) (Fig. 3c and Supplementary Fig. 9b). Glu-microtubules do not significantly grow or shrink over a period of minutes<sup>20</sup>, in agreement with the dynamics of the RNA granules we observe.

APC is an unusual plus-end tracking protein that associates with only a minority of microtubules, and it has been observed in particular at the plus ends of Glu-microtubules<sup>21</sup>. In migrating cells APC is attached to a subset of the microtubules growing into protrusions towards the leading edge<sup>22,23</sup>. When NIH/3T3 cells were induced to



**Figure 3 | Localized RNA granules are anchored at the plus ends of detyrosinated microtubules.** **a**, Cells were co-transfected with MS2-GFP and the  $\beta$ -globin-24bs/*Pkp4* RNA. Fluorescence intensity in granules at protrusive areas (left panel, red circle) or within the cytoplasm (right panel, red circle) was monitored before and after photobleaching. The arrow indicates time of bleach. Curves represent average values of ten and five independent experiments, respectively. Error bars indicate s.d. **b, c**, Confocal fluorescence images of cells expressing MS2-GFP, the  $\beta$ -globin-24bs/*Pkp4* RNA and either RFP-tubulin (**b**) or fixed and stained with anti-Glu-tubulin antibody (**c**). Panels show edges of protrusive areas. Scale bar, 3  $\mu$ m. **d**, Cells expressing mRFP, MS2-GFP and the  $\beta$ -globin-24bs/*Pkp4* RNA were treated with nocodazole (Noc.) or cytochalasin D (Cyt.D). The percentage of cells exhibiting localized RNA distribution was quantified as in Fig. 2d,  $n = 3$ ; asterisk,  $P$ -value < 0.005 by two-tailed  $t$ -test versus untreated control (-). Error bars,  $\pm 1$  s.d. **e**, Enrichment of the indicated mRNAs in pseudopodia was determined in cells extending protrusions in response to LPA in the absence or presence of nocodazole. Error bars indicate s.e.m.,  $n = 3$ ; asterisk,  $P$ -value < 0.005 by paired, two-tailed  $t$ -test versus untreated control.



**Figure 4 | APC associates with RNP complexes containing FMRP and is required for localization of RNAs in protrusions.** **a**, Confocal fluorescence images of cells expressing MS2-GFP, the  $\beta$ -globin-24bs/*Pkp4* RNA and 3xOFP-APC. Panels show edges of protrusive areas. **b**, NIH/3T3 lysates were immunoprecipitated (IP) with control antibody (IgG), anti-APC antibody (anti-APC) or no antibody (-) and analysed to detect the indicated proteins and mRNAs. **c**, Same as in **b**, except that before immunoprecipitation lysates were treated (+) or not (-) with RNase. **d**, Lysates of cells transfected with Flag-FMRP or vector were immunoprecipitated with anti-Flag antibody and analysed to detect the indicated proteins and mRNAs. **e**, Confocal fluorescence images of cells expressing MS2-GFP, the  $\beta$ -globin-24bs/*Pkp4* RNA and RFP-FMRP. Panels show edges of protrusive areas. **f**, Cells were transfected with shRNAs against luciferase (shLuc) or *Apc* (shAPC), individually or in combination as indicated. Left panel, normalized *Apc* mRNA levels; right panel, enrichment of the indicated mRNAs in pseudopodia in response to LPA ( $n = 2-3$ ); asterisk,  $P$ -value < 0.05 by paired, two-tailed  $t$ -test versus shLuc control. **g**, shLuc- and shAPC-expressing cells were transfected with MS2-GFP and the  $\beta$ -globin-24bs/*Pkp4* RNA. The percentage of cells exhibiting localized RNA distribution, as described in Fig. 2d, was quantified ( $n = 7$ ); asterisk,  $P$ -value < 0.0001 by two-tailed  $t$ -test. All error bars indicate s.e.m.; all scale bars are 5  $\mu$ m.

spread on a fibronectin-coated surface, we found that GFP-APC was mainly concentrated in granules at the tips of protrusions, reminiscent of the distribution exhibited by the localized RNAs (Supplementary Fig. 10a, arrows). Indeed, co-expression of a localized reporter RNA with APC tagged with three copies of orange fluorescent protein (3×OFP-APC) revealed that >90% ( $n = 80$ ) of the RNA granules at the tips of protrusive areas co-localize with APC (Fig. 4a, arrows and Supplementary Fig. 10b). Furthermore, FRAP analysis showed that APC present in granules is stably anchored there, exhibiting a very slow exchange rate (Supplementary Fig. 8b and Supplementary Movie 5). Taken together, these data strongly suggest that APC and the localized RNAs are part of the same complex.

To test this hypothesis, we asked if these RNAs bind specifically to endogenous APC. Indeed, APC co-immunoprecipitated with both *Rab13* and *Pkp4* mRNAs (Fig. 4b and Supplementary Fig. 11a). Importantly, under high stringency conditions, APC associated preferentially with RNAs enriched in pseudopodia (Supplementary Fig. 11b, c). Moreover, APC also bound the cytoplasmic poly(A)-binding protein (PABP1). This association was specific as neither Ran (Supplementary Fig. 11a) nor importin- $\beta$  (data not shown) was detected. Furthermore, interaction of APC with PABP1 was disrupted by pre-treatment with RNase, indicating that it is mediated through RNA (Fig. 4b, c). We also found that APC associates with FMRP (Fig. 4b), a known translational regulator of localized RNAs in other systems<sup>24,25</sup>. FMRP co-localized with the RNA granules at tips of protrusions and co-precipitated with endogenous localized mRNAs (Fig. 4d, e). We conclude that APC is a component of RNP complexes that contain localized RNAs, PABP1 and FMRP.

To test directly whether APC mediates anchoring of the localized RNAs, we knocked down APC expression using short-hairpin RNAs (shRNAs) (Fig. 4f and Supplementary Fig. 12a, b). Knockdown of APC did not significantly affect the ability of cells to extend protrusions (see Supplementary Figs 13 and 14, and data not shown), but reduced both the enrichment of the endogenous *Rab13* and *Pkp4* transcripts in pseudopodia as well as the localization of the MS2 reporter RNA in protrusions (Fig. 4f, g). It is unlikely that this effect of APC on RNA localization is mediated indirectly through the effects of APC on transcription or microtubule dynamics. APC knockdown does not affect the steady-state levels of localized RNAs (data not shown) and, consistent with previous reports<sup>26,27</sup>, it does not affect the overall microtubule organization or the presence of Glu-microtubules (Supplementary Fig. 13). APC knockdown did affect the organization of acetylated microtubules, causing more cells to exhibit short acetylated microtubules (Supplementary Figs 14 and ref. 26). However, RNA granules are not associated with acetylated microtubules (Supplementary Fig. 9a) and can form in cells with short acetylated microtubules (Supplementary Fig. 9a, middle panels). Thus, we conclude that APC is directly required for accumulation and anchoring of RNAs in protrusions.

This study provides the first genome-wide identification of asymmetrically distributed RNAs in fibroblasts. We show that, in response to migratory stimuli, >50 mRNAs accumulate in cellular protrusions of fibroblasts, revealing that even in less differentiated polarized cells, RNA localization mechanisms are widely used. Several RNAs have also been recently reported to associate *in vitro* with mitotic microtubules, but the mechanism is unknown<sup>28</sup>. Our data suggest a novel anchoring mechanism in which specific RNAs accumulate in stable granules at the plus ends of Glu-microtubules. The tumour suppressor protein APC has an essential role in this plus-end anchoring mechanism, and associates both with RNA-binding proteins and with specific RNAs. The disruption or loss of function of APC affects both polarization and cell migration<sup>26,29,30</sup>, and we speculate that at least some of the effects of this tumour suppressor are mediated through its ability to anchor mRNAs at the tips of cellular protrusions.

## METHODS SUMMARY

**Pseudopodia and cell body isolation.** To isolate pseudopodia and cell bodies in response to LPA, we followed the protocol described by ref. 13 with some modifications. Specifically, LPA was added in the bottom chamber for 1 h and the cells were fixed with 0.3% methanol-free formaldehyde (Polysciences, Inc.). Pseudopodia and cell bodies were scraped into crosslink reversal buffer (100 mM Tris pH 6.8, 5 mM EDTA, 10 mM dithiothreitol and 1% SDS), the extracts were incubated at 70 °C for 45 min and used for protein and RNA isolation. To isolate pseudopodia and cell bodies in response to fibronectin, cells were placed in serum-free media in the upper compartment of a Transwell insert equipped with a 3- $\mu$ m porous polycarbonate membrane whose underside only was coated with 5  $\mu$ g ml<sup>-1</sup> fibronectin. Cells were allowed to extend pseudopodial protrusions for 1 h and were subsequently fixed and processed as described above.

**RNA isolation and analysis.** RNA was isolated using Trizol LS reagent (Invitrogen) and contaminating DNA was removed by treatment with RQ1 DNase (Promega) for 30 min at 37 °C. One microgram of RNA was reverse transcribed using SuperscriptII Reverse Transcriptase (Invitrogen) and random hexamer primers, according to the manufacturer's instructions. cDNA was used for PCR reactions in a GeneAmp PCR System 9700 (Applied Biosystems).

**Microarray analysis.** Biotin-labelled cRNAs were generated from total RNA of pseudopodia and cell body fractions and were hybridized to GeneChip Mouse Genome 430 2.0 arrays (Affymetrix). Details about the RNA preparation and hybridization protocols can be found at <http://www.healthsystem.virginia.edu/internet/biomolec/microarray.cfm>. Results were analysed using the GeneChip Operating Software (GCOS) platform or the dChip software (<http://www.dchip.org>). Complete lists of RNAs significantly enriched in pseudopodia or cell bodies are presented in Supplementary Tables 2 and 3.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 December 2007; accepted 5 March 2008.

1. St Johnston, D. Moving messages: the intracellular localization of mRNAs. *Nature Rev. Mol. Cell Biol.* **6**, 363–375 (2005).
2. Forrest, K. M. & Gavis, E. R. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. *Curr. Biol.* **13**, 1159–1168 (2003).
3. Babu, K., Cai, Y., Bahri, S., Yang, X. & Chia, W. Roles of Bifocal, Homer, and F-actin in anchoring Oskar to the posterior cortex of *Drosophila* oocytes. *Genes Dev.* **18**, 138–143 (2004).
4. Erdelyi, M., Michon, A. M., Guichet, A., Glotzer, J. B. & Ephrussi, A. Requirement for *Drosophila* cytoplasmic tropomyosin in oskar mRNA localization. *Nature* **377**, 524–527 (1995).
5. Jankovics, F., Sinka, R., Lukacsovich, T. & Erdelyi, M. MOESIN crosslinks actin and cell membrane in *Drosophila* oocytes and is required for OSKAR anchoring. *Curr. Biol.* **12**, 2060–2065 (2002).
6. Beach, D. L., Salmon, E. D. & Bloom, K. Localization and anchoring of mRNA in budding yeast. *Curr. Biol.* **9**, 569–578 (1999).
7. Liu, G. *et al.* Interactions of elongation factor 1 $\alpha$  with F-actin and  $\beta$ -actin mRNA: implications for anchoring mRNA in cell protrusions. *Mol. Biol. Cell* **13**, 579–592 (2002).
8. Alarcon, V. B. & Elinson, R. P. RNA anchoring in the vegetal cortex of the *Xenopus* oocyte. *J. Cell Sci.* **114**, 1731–1741 (2001).
9. Delanoue, R. & Davis, I. Dynein anchors its mRNA cargo after apical transport in the *Drosophila* blastoderm embryo. *Cell* **122**, 97–106 (2005).
10. Condeelis, J. & Singer, R. H. How and why does  $\beta$ -actin mRNA target? *Biol. Cell* **97**, 97–110 (2005).
11. Leung, K. M. *et al.* Asymmetrical  $\beta$ -actin mRNA translation in growth cones mediates attractive turning to netrin-1. *Nature Neurosci.* **9**, 1247–1256 (2006).
12. Mingle, L. A. *et al.* Localization of all seven messenger RNAs for the actin-polymerization nucleator Arp2/3 complex in the protrusions of fibroblasts. *J. Cell Sci.* **118**, 2425–2433 (2005).
13. Cho, S. Y. & Klemke, R. L. Purification of pseudopodia from polarized cells reveals redistribution and activation of Rac through assembly of a CAS/Crk scaffold. *J. Cell Biol.* **156**, 725–736 (2002).
14. Moissoglu, K. & Schwartz, M. A. Integrin signalling in directed cell migration. *Biol. Cell* **98**, 547–555 (2006).
15. Vikesaa, J. *et al.* RNA-binding IMPs promote cell adhesion and invadopodia formation. *EMBO J.* **25**, 1456–1468 (2006).
16. Jambhekar, A. & Derisi, J. L. Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA* **13**, 625–642 (2007).
17. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
18. Galjart, N. CLIPs and CLASPs and cellular dynamics. *Nature Rev. Mol. Cell Biol.* **6**, 487–498 (2005).
19. Akhmanova, A. & Hoogenraad, C. C. Microtubule plus-end-tracking proteins: mechanisms and functions. *Curr. Opin. Cell Biol.* **17**, 47–54 (2005).

20. Infante, A. S., Stein, M. S., Zhai, Y., Borisy, G. G. & Gundersen, G. G. Detyrosinated (Glu) microtubules are stabilized by an ATP-sensitive plus-end cap. *J. Cell Sci.* **113**, 3907–3919 (2000).
21. Wen, Y. *et al.* EB1 and APC bind to mDia to stabilize microtubules downstream of Rho and promote cell migration. *Nature Cell Biol.* **6**, 820–830 (2004).
22. Mimori-Kiyosue, Y., Shiina, N. & Tsukita, S. Adenomatous polyposis coli (APC) protein moves along microtubules and concentrates at their growing ends in epithelial cells. *J. Cell Biol.* **148**, 505–518 (2000).
23. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. H. & Nelson, W. J. The adenomatous polyposis coli tumor suppressor protein localizes to plasma membrane sites involved in active cell migration. *J. Cell Biol.* **134**, 165–179 (1996).
24. Bagni, C. & Greenough, W. T. From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nature Rev. Neurosci.* **6**, 376–387 (2005).
25. Jin, P., Alisch, R. S. & Warren, S. T. RNA and microRNAs in fragile X mental retardation. *Nature Cell Biol.* **6**, 1048–1053 (2004).
26. Kroboth, K. *et al.* Lack of adenomatous polyposis coli protein correlates with a decrease in cell migration and overall changes in microtubule stability. *Mol. Biol. Cell* **18**, 910–918 (2007).
27. Kita, K., Wittmann, T., Nathke, I. S. & Waterman-Storer, C. M. Adenomatous polyposis coli on microtubule plus ends in cell extensions can promote microtubule net growth with or without EB1. *Mol. Biol. Cell* **17**, 2331–2345 (2006).
28. Blower, M. D., Feric, E., Weis, K. & Heald, R. Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. *J. Cell Biol.* **179**, 1365–1373 (2007).
29. Watanabe, T. *et al.* Interaction with IQGAP1 links APC to Rac1, Cdc42, and actin filaments during cell polarization and migration. *Dev. Cell* **7**, 871–883 (2004).
30. Etienne-Manneville, S. & Hall, A. Cdc42 regulates GSK-3 $\beta$  and adenomatous polyposis coli to control cell polarity. *Nature* **421**, 753–756 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank Y. Bao for the microarray and real-time PCR analysis, and J. Lykke-Andersen, I. Nathke, J. T. Parsons, R. Darnell and G. Gundersen for plasmids and antibodies. S.M. is a fellow of the Leukemia and Lymphoma Society. This work was supported by a grant from the NIH to I.G.M., and by the James and Rebecca Craig Foundation.

**Author Contributions** S.M. performed the experiments. S.M., K.M. and I.G.M. designed the experiments and analysed the data. K.M. provided reagents. S.M. and I.G.M. wrote the manuscript.

**Author Information** The microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE10230. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to I.G.M. ([igm9c@virginia.edu](mailto:igm9c@virginia.edu)) or S.M. ([sm2ju@virginia.edu](mailto:sm2ju@virginia.edu)).



## METHODS

**Plasmid constructs.** A plasmid encoding the human  $\beta$ -globin mRNA was provided by J. Lykke-Andersen. The plasmid contains all the coding exons and introns of human  $\beta$ -globin, 54 nucleotides of the  $\beta$ -globin 5' UTR, 69 nucleotides of the  $\beta$ -globin 3' UTR followed by 24 repeats of the MS2-binding site cloned into the HindIII, XbaI sites of pcDNA3 (Invitrogen). To facilitate cloning, a multiple cloning site with sequence 5'-ATCGATGGTACCGCTACGGATA-TCTCTGAG-3' was introduced between the XbaI and ApaI sites, generating plasmid  $\beta$ -globin-24bs/-. Various 3' UTRs were PCR amplified from mouse genomic DNA (BD Biosciences Clontech) with primers carrying appropriate restriction sites and were ligated into the NheI and XhoI sites (for the *Pkp4* and *Rab13* 3' UTRs) or into the XbaI and ApaI sites (for the *Rac1* and *RhoA* 3' UTRs) of the  $\beta$ -globin-24bs/- plasmid, to generate plasmids  $\beta$ -globin-24bs/*Pkp4*,  $\beta$ -globin-24bs/*Rab13*,  $\beta$ -globin-24bs/*Rac1* and  $\beta$ -globin-24bs/*RhoA*. The exact 3' UTRs correspond to: nucleotides 3858–4612 of NM\_026361 (*Pkp4* UTR), nucleotides 91295–92092 of NT\_039254 (*Rab13* UTR), nucleotides 776–2240 of NM\_009007 (*Rac1* UTR) and nucleotides 1009–2071 of NM\_016802 (*RhoA* UTR). To generate plasmids  $\beta$ -globin-0bs/*Pkp4* and  $\beta$ -globin-0bs/*Rab13*, plasmids  $\beta$ -globin-24bs/*Pkp4* and  $\beta$ -globin-24bs/*Rab13*, respectively, were digested with NotI and NheI (to remove the fragment containing the 24 MS2-binding sites), the ends were blunted with T4 DNA polymerase and re-ligated. To generate constructs *Rab13/Rab13* and *Rab13/RhoA*, the genomic sequence containing the coding exons and intervening introns of the *Rab13* gene was PCR amplified from mouse genomic DNA with primers introducing an EcoRI site at the 5' end and a KpnI site at the 3' end. A Flag tag was ligated at the EcoRI site in frame with the *Rab13* ORF. The *Rab13* 3' UTR or the *RhoA* 3' UTR were ligated at the KpnI site through blunt-end ligation. The two resulting fragments were ligated into pEGFP-C1 vector from which the GFP sequence had been previously removed.

The plasmid pcNMS2, containing an oligomerization-defective MS2 coat protein mutant, was provided by J. Lykke-Andersen. The GFP sequence was PCR amplified from plasmid pEGFP-N3 (Clontech) with primers introducing BamHI and XhoI sites and was ligated at the XhoI site with oligonucleotides containing the SV40 large T-Ag NLS. The GFP-NLS fragment was ligated into the BamHI and NotI sites of pcNMS2 to generate the pcNMS2-GFP-NLS plasmid expressing the MS2-GFP protein. RFP-tubulin was generated from pEGFP-tubulin (Clontech) by replacing the NheI/XhoI GFP fragment with a PCR-amplified fragment of mRFP. To generate EB1-GFP and EB1-RFP expressing plasmids, the mouse EB1 cDNA was amplified by RT-PCR from total NIH/3T3 RNA with primers introducing a KpnI site at the 5' end and a BglII site at the 3' end. The EB1 fragment was either ligated into KpnI/BamHI sites of pEGFP-N3 or into KpnI/XhoI sites of pcDNA3 (Invitrogen) together with a BglII/XhoI PCR-amplified fragment of mRFP. The plasmid expressing GFP-APC was provided by I. Nathke. The SacI/BamHI APC fragment was ligated with a NheI/SacI fragment containing three copies of OFP (generated through sequential PCR amplification and ligation of individual OFP fragments) into the XbaI and BamHI sites of pKH3 to generate plasmid 3 $\times$ OFP-APC. To generate RFP-FMRP, a plasmid expressing GFP-FMRP was used (provided by R. Darnell) and the NheI/SacI fragment, encoding GFP, was replaced with a NheI/SacI PCR-amplified fragment of mRFP. To generate Flag-FMRP, the SacI/EcoRI fragment encoding FMRP (from the RFP-FMRP plasmid) was ligated together with a NheI/SacI fragment encoding the Flag tag into the XbaI/EcoRI sites of pKH3. To generate RFP-DCP1B, a plasmid expressing Flag-DCP1B was used (provided by J. Lykke-Andersen). The BamHI (blunted with T4 DNA polymerase)/NotI fragment encoding human DCP1B was ligated with a HindIII/NotI (blunted with T4 DNA polymerase) fragment of mRFP into HindIII and NotI sites of pcDNA3.

For knockdown experiments, oligonucleotides were synthesized targeting different regions of the mouse *Apc* mRNA. Sequences of the oligonucleotides are as follows: shAPC 2 sense oligonucleotide, 5'-GATCCCCGAATCAACCAGGC-ATAATATTCAAGAGATATTATGCCTGGTTGATTCTTTTGGAAA-3' and shAPC 2 antisense oligonucleotide, 5'-AGCTTTTCCAAAAAGAATCAACCA-GGCATAATATCTCTTGAATATTATGCCTGGTTGATTCTGGG-3'; shAPC 4 sense oligonucleotide, 5'-GATCCCCCTAAGTGATCTGACAATAGATTCAAG-AGATCTATTGTCAGATCACTTATTTTGGAAA-3' and shAPC 4 antisense oligonucleotide, 5'-AGCTTTTCCAAAAATAAGTGATCTGACAATAGATCTCTTGAATCTATTGTCAGATCACTTAGGG-3'; shAPC 5 sense oligonucleotide, 5'-GATCCCCCACTACAGTGAACGTTATTCTCAAGAGAATAACG-TTCACTGTAGTTGTTTTGGAAA-3' and shAPC 5 antisense oligonucleotide, 5'-AGCTTTTCCAAAACACTACAGTGAACGTTATTCTCTTGAATAA-CGTTCACTGTAGTTGGGG-3'. Bold characters indicate the *Apc* mRNA targeting sequence; underlined characters indicate the 9-bp hairpin loop. Sequences of control oligonucleotides targeting luciferase were: sense

oligonucleotide, 5'-GATCCCCGTACGCGGAATACTTCGATTCAAGAGA-TCGAAGTATTCGCGGTACGTTTTTGGAAA-3', antisense oligonucleotide, 5'-AGCTTTTCCAAAAACGTACGCGGAATACTTCGATTCTTGAATCGA-AGTATTCGCGGTACGCGG-3'. The sense and antisense oligonucleotides were annealed and ligated into the BglII and HindIII sites of the pSuper vector.

**Cell culture and transfection.** NIH/3T3 cells were grown in DMEM supplemented with 10% calf serum, sodium pyruvate, penicillin and streptomycin (Invitrogen). For live cell imaging, plasmid constructs were transfected with Effectene (Qiagen) according to the manufacturer's instructions. Twenty-four hours after transfection cells were plated for ~2 h on Lab-Tek coverglass chambers (Nalge nunc International) coated with 5  $\mu$ g ml<sup>-1</sup> fibronectin and fluorescence was visualized by confocal microscopy. Where indicated, cells were treated at 37 °C for 30 min with 10  $\mu$ M nocodazole or 1  $\mu$ M cytochalasin D. For pseudopodia/cell body fractionation or for shRNA-mediated knockdown experiments, cells were electroporated using the Gene Pulser II electroporation system (Bio-Rad) with plasmid constructs (5  $\mu$ g per 6  $\times$  10<sup>6</sup> cells) or with pSuper constructs encoding shRNAs (40  $\mu$ g per 6  $\times$  10<sup>6</sup> cells), respectively. Electroporation efficiency, based on co-transfected GFP, was generally >70%. Electroporated cells were processed after 72 h.

**Pseudopodia and cell body isolation.** To isolate pseudopodia and cell bodies of cells induced to migrate with LPA, we followed the protocol described by ref. 13 with some modifications. Cells were serum-starved overnight and 1.5  $\times$  10<sup>6</sup> cells were placed in the upper compartment of a Transwell insert (24-mm diameter, Costar) equipped with a 3- $\mu$ m porous polycarbonate membrane coated on both sides with 5  $\mu$ g ml<sup>-1</sup> fibronectin. Cells were allowed to spread on the upper surface of the membrane for 2 h. LPA (150 ng ml<sup>-1</sup>) was then added in the bottom chamber to induce the cells to extend pseudopodial protrusions. After 1 h the cells were briefly rinsed with PBS and fixed with 0.3% methanol-free formaldehyde (Polysciences, Inc.) in PBS for 10 min at room temperature. Glycine was added to 250 mM for 5 min at room temperature and the cells were washed twice with PBS. To isolate pseudopodia, cell bodies on the upper membrane surface were manually removed with a cotton swab and laboratory paper and pseudopodia on the underside of the membrane were scraped into crosslink reversal buffer (100 mM Tris pH 6.8, 5 mM EDTA, 10 mM dithiothreitol and 1% SDS). Cell bodies were similarly isolated except that pseudopodia on the underside of the membrane were manually removed and cell bodies were scraped into crosslink reversal buffer. Extracts were incubated at 70 °C for 45 min to reverse the formaldehyde-induced crosslinks. Proteins were directly analysed by western blot otherwise total RNA was isolated and processed as described below.

For fractionation after nocodazole treatment, cells were processed as above except that nocodazole was added to 10  $\mu$ M during the last 25 min of the assay.

To isolate pseudopodia and cell bodies of cells migrating towards fibronectin, ~1  $\times$  10<sup>6</sup> cells were placed in serum-free media in the upper compartment of a Transwell insert equipped with a 3- $\mu$ m porous polycarbonate membrane whose underside only was coated with 5  $\mu$ g ml<sup>-1</sup> fibronectin. Cells were allowed to extend pseudopodial protrusions towards the fibronectin-coated surface for 1 h and were subsequently fixed and processed as described above.

**RNA isolation and analysis.** Total RNA or RNA from pseudopodia and cell body fractions was isolated using Trizol LS reagent (Invitrogen) and contaminating DNA was removed by treatment with RQ1 DNase (Promega) for 30 min at 37 °C. One microgram of RNA was reverse transcribed using SuperscriptII Reverse Transcriptase (Invitrogen) and random hexamer primers according to the manufacturer's instructions. cDNA was used for PCR reactions in a GeneAmp PCR System 9700 (Applied Biosystems) to detect different RNAs using the following primer pairs: *Rab13* F 5'-GCCTACCAAGTGTGGCTCTT-3', *Rab13* R 5'-TCCACGTAATAGGCGGTAG-3', *Rab13* UTR F 5'-AGGCT-GCTAGCGAGCATTTCTTGCCTCTAT-3', *Rab13* UTR R 5'-AATGGCTC-GAGCCATTCATTTCTTCTTCC-3'; *Pkp4* F 5'-AGGCTGCTAGCCAGGGAA-GTGAGGAAACC-3', *Pkp4* R 5'-AATGGCTCGAGAAACATGAAGGGCA-TCC-3'; *Ankr25* F 5'-TTCAAAGCCAGAAAGCAAG-3', *Ankr25* R 5'-AGGTGACAAAGGGTGGTGAG-3'; *Inpp1*\* F 5'-TTTGAAGTGAATGGG-GATAAC-3', *Inpp1*\* R 5'-AATAGTCAGATAGTCAAACTCATGG-3' (these primers detect a likely RNA isoform of *Inpp1* with a longer 3' UTR); *Apc* F 5'-CCTCTCACCGGAGTAAGCAG-3', *Apc* R 5'-GTCGTCTGGGAGGTAT-GAA-3'; *Arpc3* F 5'-CACGGACATTGTGGATGAAG-3', *Arpc3* R 5'-CCACC-ACTTGCTGGCTTTAT-3'; Flag *Rab13* F 5'-TAATACGACTCACTATAGGG-3', Flag *Rab13* R 5'-TCCACGTAATAGGCGGTAG-3';  $\beta$ -globin F 5'-TTGA-GTCCTTTGGGGATCTG-3',  $\beta$ -globin R 5'-CACTGGTGGGGTGAATCTT-3';  $\beta$ -actin F 5'-TGTTACCAACTGGGACGACA-3',  $\beta$ -actin R 5'-GCTGTGGT-GGTGAAGCTGTA-3'.

The identity of the amplified products was verified either by sequencing or in the case of transfected RNAs by ensuring that no product was amplified when using RNA from vector-transfected cells. PCR reactions contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM each dNTPs, 0.5  $\mu$ M each

primers and 2.5 U Taq polymerase. The amount of cDNA and the number of cycles were varied for each primer pair, to ensure amplification was within the linear phase. This was verified in all experiments by including decreasing amounts of selected samples.

To calculate the enrichment of RNAs in pseudopodia, equal amounts of RNA from pseudopodia and cell body fractions were analysed by RT-PCR, the signals were quantified using WCIF ImageJ software, normalized to the control *Arpc3* mRNA and enrichment in pseudopodia was defined as (pseudopodia signal/cell body signal) - 1, when pseudopodia signal > cell body signal or as (-1) × [(cell body signal/pseudopodia signal) - 1], when pseudopodia signal < cell body signal.

**Microarray analysis.** Microarray analysis was performed at the Biomolecular Research Facility at the University of Virginia. Total RNA from pseudopodia and cell body fractions was analysed on an Agilent BioAnalyser and was deemed to be of good quality. Biotin-labelled cRNAs were generated and hybridized to GeneChip Mouse Genome 430 2.0 arrays (Affymetrix), which contain 45,000 probe sets analysing the expression level of over 39,000 transcripts and variants from over 34,000 well-characterized mouse genes. Details about the RNA preparation and hybridization protocols can be found at <http://www.healthsystem.virginia.edu/internet/biomolec/microarray.cfm>.

In all hybridization experiments, quality assessment variables, such as background, noise, GAPDH 3'/5' ratio, were within the acceptance limits.

For experiments analysing pseudopodia and cell body fractions from cells induced with LPA, results were analysed using the GeneChip Operating Software (GCOS) platform. RNAs were considered to be significantly enriched in pseudopodia if the *P*-value was significant, the fold change in signal intensity was greater than 2.2 and the absolute difference in signal intensity was greater than 100.

For experiments analysing pseudopodia and cell body fractions from cells migrating towards fibronectin, results were analysed using the dChip software (<http://www.dchip.org>). RNAs were considered to be significantly enriched in pseudopodia if the *P*-value was less than 0.05, the fold change in signal intensity was greater than 1.5 and the absolute difference in signal intensity was greater than 10 times the noise level.

**Western blot, immunofluorescence and immunoprecipitation.** For western blot and immunofluorescence, the following antibodies were used: mouse monoclonal anti-Ran (Transduction Laboratories), rabbit anti-pY397-FAK (Biosource International), monoclonal anti-FAK (BD Transduction Labs),

anti-APC (C-20) (Santa Cruz Biotechnology), rabbit anti-PABP1 (Cell Signaling Technology), anti-FMRP (clone 1C3, Chemicon), anti-acetylated tubulin (Clone 6-11B-1, Sigma), anti- $\alpha$ -tubulin (clone DM1A, Sigma), anti-Glu-tubulin (provided by G. Gundersen).

For immunoprecipitation, anti-APC (C-20) or control antibody (rabbit anti-HA tag, Santa Cruz Biotechnology) was bound on protein-A beads, otherwise anti-Flag M2 agarose beads (Sigma) were used. In all cases, antibody-bound beads were pre-incubated with 200  $\mu\text{g ml}^{-1}$  *Escherichia coli* tRNA, 200  $\mu\text{g ml}^{-1}$  herring sperm DNA, 200  $\mu\text{g ml}^{-1}$  RNase-free BSA and 50  $\mu\text{g ml}^{-1}$  glycogen. Cells spreading on fibronectin-coated plates for approximately 2 h were lysed in lysis buffer (10 mM Tris pH 7.5, 100 mM NaCl, 2.5 mM  $\text{MgCl}_2$ , 0.5% Triton X-100) supplemented with protease inhibitors (leupeptin, pepstatin, aprotinin) and RNase inhibitor (0.5 U  $\mu\text{l}^{-1}$ ). Lysates were centrifuged at 10,000g for 10 min and incubated with antibody-bound beads at 4 °C for 3 h. Beads were washed five times either with lysis buffer or, where indicated, with high stringency wash buffer (50 mM Tris pH 7.5, 500 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS). The immunoprecipitated material was released by incubation in lysis buffer containing 1% SDS. A fraction of the released material was analysed by western blot. The remainder was used for total RNA isolation and RT-PCR analysis. For RNase treatment, 1 mM  $\text{CaCl}_2$  was added in the lysis buffer and after the initial centrifugation, RNase A (30  $\mu\text{g ml}^{-1}$ ) and micrococcal nuclease (70 U  $\text{ml}^{-1}$ ) were added, the lysate was incubated at 30 °C for 10 min and subsequently centrifuged at 10,000g for 5 min. The supernatant was used in immunoprecipitations as described above.

**Confocal microscopy and photobleaching.** Imaging and photobleaching were performed with a Zeiss LSM 510 Meta confocal microscope operated by LSM-FCS software (Carl Zeiss). Temperature was maintained at 37 °C. Medium pH was controlled by addition of 25 mM HEPES buffer. EGFP was excited with the 488-nm line of an argon laser at 30% power, 5% transmission. mRFP, OFP and DsRed were excited with the 543-nm line of a HeNe laser at 80% transmission. For photobleaching, regions of interest were selected and bleached with the 488-nm laser line at 100% power, 100% transmission for two iterations. Fluorescence recovery within the region of interest was monitored for ~60 s. For each experiment, three images were recorded pre-bleach. Mean intensities in the bleached area were measured, background signal was subtracted, intensities were corrected for bleaching during imaging and were expressed as a percentage of the pre-bleach intensity.

## LETTERS

## Life without RNase P

Lennart Randau<sup>1</sup>, Imke Schröder<sup>3</sup> & Dieter Söll<sup>1,2</sup>

The universality of ribonuclease P (RNase P), the ribonucleoprotein essential for transfer RNA (tRNA) 5' maturation<sup>1,2</sup>, is challenged in the archaeon *Nanoarchaeum equitans*. Neither extensive computational analysis of the genome nor biochemical tests in cell extracts revealed the existence of this enzyme. Here we show that the conserved placement of its tRNA gene promoters allows the synthesis of leaderless tRNAs, whose presence was verified by the observation of 5' triphosphorylated mature tRNA species. Initiation of tRNA gene transcription requires a purine, which coincides with the finding that tRNAs with a cytosine in position 1 display unusually extended 5' termini with an extra purine residue. These tRNAs were shown to be substrates for their cognate aminoacyl-tRNA synthetases. These findings demonstrate how nature can cope with the loss of the universal and supposedly ancient RNase P through genomic rearrangement at tRNA genes under the pressure of genome condensation.

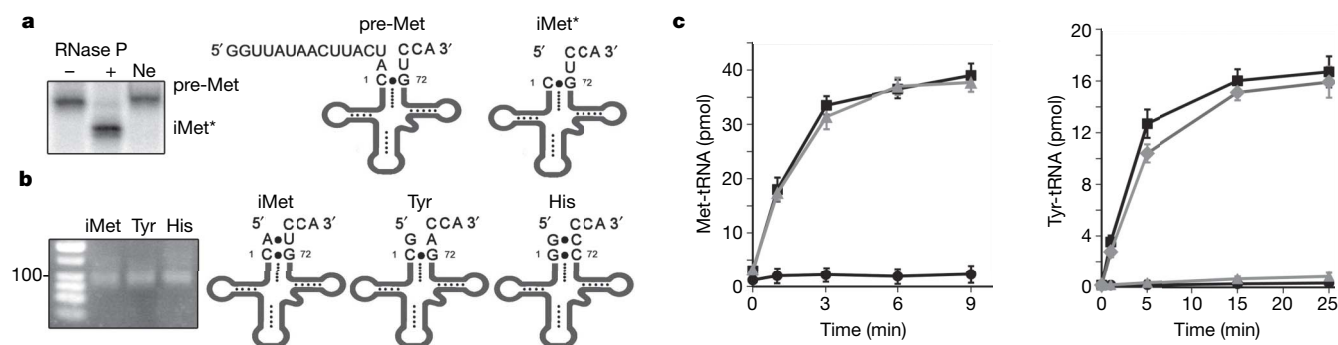
The transcription of tRNA genes generates precursor tRNA molecules with extended 5' and 3' termini that are processed into mature tRNA<sup>3</sup>. The maturation of the 5' terminus of a pre-tRNA relies solely on the endonucleolytic activity of the ubiquitous and essential ribonuclease P (RNase P)<sup>1,2</sup>. Both the substrate-binding domain and the active site are located in the RNA moiety of RNase P. Additional protein subunits serve as cofactors required for *in vivo* activity whereas *in vitro* cleavage was shown to be catalysed by the RNase P RNA molecule alone in all three domains of life<sup>4–6</sup>. The number of required protein cofactors varies from one in Bacteria to up to ten subunits in Eukarya; four to five homologues of the eukaryal RNase P proteins exist in Archaea<sup>7</sup>.

Surprisingly, analysis of genome sequence failed to reveal genes for RNase P in *N. equitans*, several species of *Pyrobaculum* and the

bacterium *Aquifex aeolicus*<sup>8</sup>. Although previous studies indicated that the missing bacterial enzyme is possibly replaced by an alternative activity reminiscent of bacterial RNase E-like enzymes<sup>9</sup>, no explanation has been found for the obvious lack of RNase P in these Archaea. Here, we focus on the analysis of the absence of RNase P in *N. equitans* and the genomic rearrangements necessary to cope with its loss.

First we used a computational approach to verify that no RNase P RNA is present in the genome of *N. equitans*. An algorithm was developed which, based on the observation that non-coding RNA genes are easily identified in AT-rich hyperthermophiles<sup>10</sup>, searches for GC-rich islands within the AT-rich genome of *N. equitans*. This algorithm clearly identifies all tRNAs and ribosomal RNA, as well as certain small nucleolar RNAs. However, no RNase P RNA, complete or fragmented, was discovered. Second, we investigated whether RNase P activity is detectable in cell extracts or total RNA extracts of *N. equitans*. A pre-tRNA based on the *N. equitans* initiator tRNA was produced, internally labelled in the presence of [ $\alpha$ -<sup>32</sup>P]ATP and used as substrate for RNase P cleavage assays (Fig. 1a). The 15-nucleotide leader sequence was easily released by *Escherichia coli* RNase P RNA, but no cleavage activity was observed with the cell extract or total RNA preparation of *N. equitans* (Fig. 1a). Assay conditions including time, pH, salt concentration and length of leader sequence of the pre-tRNA substrate were varied without yielding any detectable activity (data not shown).

The lack of RNase P activity raised the question of how an organism could live without RNase P and led us to investigate whether tRNAs are produced without 5' leader sequences. This hypothesis would require that transcription of all tRNA species starts at the 5' terminus of the mature tRNA. Therefore we analysed the promoter



**Figure 1 | RNase P cleavage assay and tRNA genes in *N. equitans*.** **a**, RNase P cleavage assay. The internally radioactively labelled pre-tRNA<sub>i</sub><sup>Met</sup> transcript (pre-Met) is cleaved in the presence of *E. coli* RNase P RNA (+) but not in the presence of cell extract of *N. equitans* (Ne). The cleavage product (iMet\*) was isolated and sequenced. **b**, *N. equitans* tRNA sequences. The RT-PCR products were separated on a 3% ethidium-bromide-stained agarose gel and compared with a PCR marker with the 100-bp fragment

indicated. The sequenced termini of the tRNAs are shown with numbering according to ref. 30. **c**, Methionylation and tyrosylation of 5' extended tRNA transcripts (filled squares) at 50 °C. Control reactions without tRNA (filled circles), 5' monophosphorylated tRNA<sup>Tyr</sup> (grey diamonds) and tRNA transcripts lacking the base at position -1 (grey triangles) are included. Values are the average from three assays; error bars, s.d.

<sup>1</sup>Department of Molecular Biophysics and Biochemistry, <sup>2</sup>Department of Chemistry, Yale University, New Haven, Connecticut 06520-8114, USA. <sup>3</sup>Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, California 90095, USA.

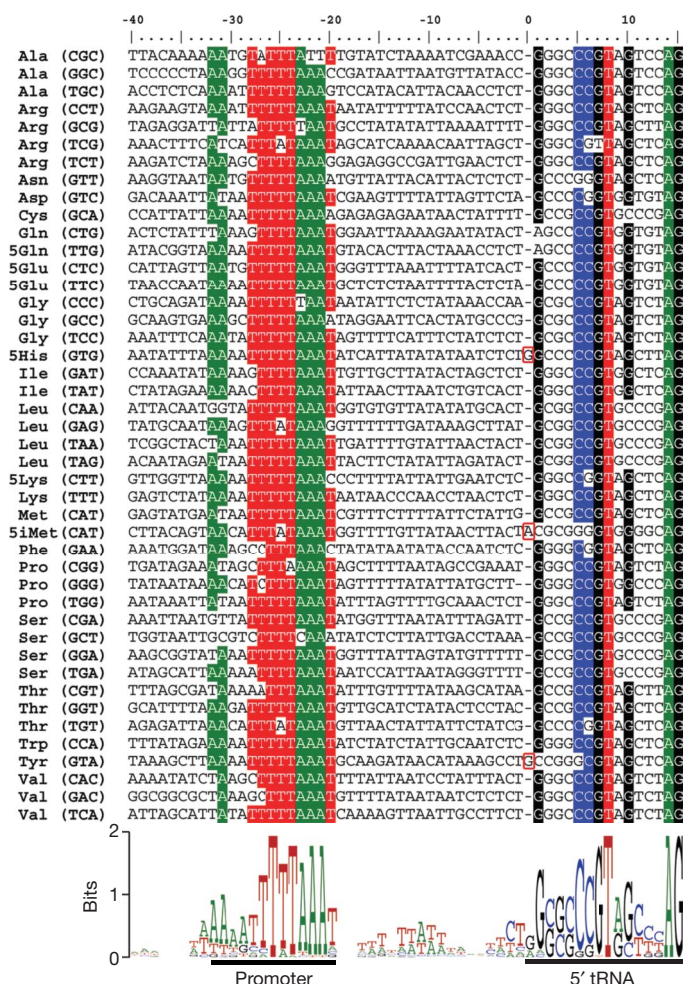


placement upstream of all tRNA (and tRNA half<sup>11</sup>) genes in the genome of *N. equitans*. It is striking that each tRNA gene possesses a potential promoter positioned 26 nucleotides upstream of the mature tRNA sequence, which was shown to be the conserved distance between promoter and transcription start site in Archaea<sup>12</sup> (Fig. 2). The strict distance of these tRNA gene promoters to the tRNA gene start allows the analysis of its architecture by a WebLogo representation<sup>13</sup>. Apart from a described hexanucleotide TTTAAA motif<sup>12</sup> that displays an invariable T at position -26, conserved adenosines are found at positions -32 and -31 and thymidine residues are conserved at positions -28, -27 and -20. A similar, but slightly less conserved, promoter placement is observed for *Pyrobaculum* species (Supplementary Fig. 1).

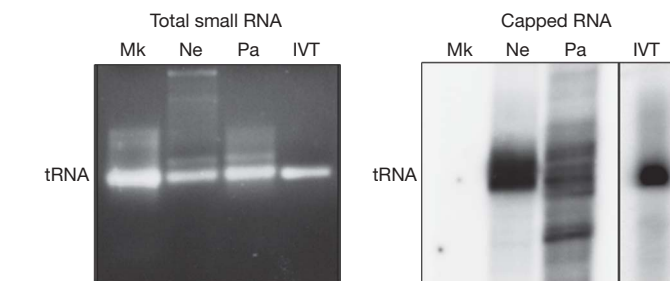
This strict promoter placement probably allows production of leaderless transcripts. Transcription is predominantly initiated with a purine (G > A) in all living organisms<sup>14</sup>. Thus G is the first nucleotide in nearly all tRNAs from these organisms; however, close examination of all tRNA genes revealed a few exceptions. Therefore we determined the sequences of the 5' termini in the tRNA population of *N. equitans*. This was accomplished by circularizing the tRNAs by ligating the 3' CCA end of the mature tRNAs to the 5' end of the same molecule. Specific tRNAs were amplified by polymerase chain reaction with reverse transcription (RT-PCR) using oligonucleotides complementary to the anticodon region of the tRNA molecule,

which allowed the subsequent sequencing of the ligation site. The first tRNA of interest was tRNA<sup>Tyr</sup>, which requires a C1–G72 base pair (bp) as the major identity element for tyrosyl-tRNA synthetase<sup>15</sup>. Sequencing revealed that the nanoarchaeal tRNA<sup>Tyr</sup> displays one additional G residue at position -1 at its 5' terminus (Fig. 1b). We confirmed that this tRNA is a substrate for the *N. equitans* TyrRS, whereas a tRNA transcript without the G-1 residue could not be tyrosylated (Fig. 1c). The phosphorylation state of the 5' end of the G-1 containing tRNA had no significant effect on aminoacylation activity. Another exception is *N. equitans* tRNA<sup>Met</sup>, which contains C1 and an extra A-1 at its 5' end (Fig. 1b). This is of special interest as all known archaeal initiator tRNAs contain an A1–U72 bp required for recognition by initiation factor aIFγ<sup>16,17</sup>. This RNA recognition element is shifted to positions -1•73 in the *N. equitans* tRNA to form an eight-bp acceptor stem (Fig. 1b). The unique elongated acceptor stem of this initiator tRNA does not affect the methionylation by the *N. equitans* methionyl-tRNA synthetase (Fig. 1c). It is pertinent to note that such an eight-bp acceptor stem would be impossible if a conventional RNase P was present, as *E. coli* RNase P catalysed cleavage between nucleotides A-1 and C1 in an *N. equitans* tRNA<sup>Met</sup> transcript (Fig. 1a). Similarly, studies on tRNA<sup>Tyr</sup> maturation indicate that an organism with RNase P would not generate tRNAs with the observed 5' purine extensions<sup>7</sup>. Finally, the nanoarchaeal tRNA<sup>His</sup> contains the conserved G-1 residue required for recognition by the histidyl-tRNA synthetase<sup>18</sup>. The universally occurring eight-bp acceptor stem of tRNA<sup>His</sup> demonstrates that the protein biosynthesis machinery is able to deal with an extra 5' nucleotide. Thus three tRNAs exist with extended 5' termini, so that transcription can initiate with a purine. This extension is in agreement with a shift of the conserved hexanucleotide promoter element from -26 to -27 nucleotides upstream of position 1 in relation to its tRNA gene (Fig. 2).

Transcription initiation at the mature 5' termini of nanoarchaeal tRNAs would result in triphosphorylated 5' termini (for example found in the initiator tRNA of *Halobacterium volcanii*<sup>19</sup>) rather than in monophosphorylated 5' termini generated by RNase P cleavage. To address whether this is the case *in vivo*, we took advantage of the specificity of the vaccinia virus capping enzyme activity. This guanylyltransferase transfers radioactively labelled GTP to a ppp-5' terminus, but not to a p-5' terminus, of an RNA molecule<sup>20</sup>. No radioactivity was transferred to the total small RNA of *Methanopyrus kandleri*, an archaeon that contains RNase P. However, labelling was observed in *N. equitans*, as witnessed by a band that co-migrates with a labelled tRNA transcript (Fig. 3). Various RNA substrates of *P. aerophilum*, including tRNA-sized molecules, were also labelled (Fig. 3). The labelled band from *N.*



**Figure 2 | Strict tRNA gene promoter placement in *N. equitans*.** The upstream region of all 44 tRNA genes (including 5' tRNA halves) was aligned. Position 1 indicates the first tRNA nucleotide. Position -1 of three tRNA species is boxed in red. The conserved promoter element is visualized by a WebLogo<sup>13</sup> in which the size of a letter indicates the degree of conservation of a nucleotide.



**Figure 3 | Detection of triphosphorylated tRNA.** Equal amounts of total small RNA from *M. kandleri* (Mk), *N. equitans* (Ne), *P. aerophilum* (Pa) and an *N. equitans* tRNA<sup>Tyr</sup> *in vitro* transcript (IVT) were separated on a 3% ethidium-bromide-stained agarose gel (left). These RNA preparations were reacted with vaccinia virus capping enzyme, the products loaded onto a denaturing 12% polyacrylamide gel, and the transfer of radioactive GTP onto triphosphorylated RNA monitored by autoradiography (right). Sequencing samples of the gel-eluted labelled RNA of *N. equitans* verified the presence of tRNA. Full-length gels are presented in Supplementary Fig. 2.

*equitans* was excised from the gel, and the RNA was eluted. RT-PCR and sequencing confirmed the presence of tRNAs. Thus, 5' triphosphorylated tRNA species are present among small RNAs extracted from *N. equitans* cells. The instability of these 5' triphosphate ends probably accounts for the fraction of 5' monophosphorylated tRNAs that could be circularized without phosphatase pretreatment, as discussed above. It has been shown that both 5' mono- and triphosphorylated suppressor tRNAs are equally active in protein biosynthesis<sup>21</sup>. This raises the question of whether both phosphorylation states of tRNA are also accepted by the *N. equitans* ribosome.

The presence of a catalytic RNA moiety and the universal conservation of RNase P imply that this enzyme is ancient, possibly a remnant of the RNA world<sup>22</sup>, and therefore present in the last universal common ancestor. An interesting question is why *N. equitans* lost RNase P. Because the *N. equitans* ancestor probably had leader-containing tRNAs and RNase P, we have to consider two driving forces of nanoarchaeal evolution that led to their absence. First, *N. equitans* has the smallest genome with the highest coding density of any sequenced genome<sup>23</sup>. It is likely that this genome reduction is a strategy of adaptation to the obligatory parasitic lifestyle of *N. equitans*<sup>24</sup>. Second, tRNA molecules are hotspots for integrative elements, as witnessed by the nearly exclusive usage of tRNA genes as attachment sites for viruses<sup>25</sup>. Clearly, an organism with leader-containing tRNAs but lacking RNase P is not viable<sup>26</sup>. We propose an intermediate situation that led to the complete loss of RNase P (Fig. 4). First, the activity of RNase P was reduced either by the loss of certain protein subunits of the ribonucleoprotein complex during genome reduction or by mutation of the RNA molecule. The fast evolutionary tempo of *N. equitans* and the integration/excision events of mobile elements at the tRNA genes allowed genome rearrangement and promoter placement that made RNase P obsolete. The organism adapted to this situation by allowing a certain ambiguity of the 5' termini of the tRNAs, as witnessed by extended and triphosphorylated tRNAs. When every tRNA gene promoter allowed the generation of leaderless tRNA transcripts, the entire RNase P molecule was lost. Consequently, this situation prevents the insertion of integrative elements in *N. equitans* and therefore the expansion of its genome. *Pyrobaculum* shares the fast evolutionary clock owing to its 'mutator' phenotype<sup>27</sup> and might represent the described intermediate state. Although most tRNA genes in *Pyrobaculum* contain a promoter in the conserved distance to the transcription start site, the promoter placement suggests the presence of small 5' leaders for a few pre-tRNAs (Supplementary Fig. 1). Furthermore, a conventional RNase P is not detectable and only one protein subunit can be identified<sup>28</sup>.

Because tRNA is the typical substrate of RNase P, it is possible that tRNA and tRNA genes predate RNase P. The emergence of RNase P would then coincide with a more flexible tRNA promoter placement or tRNA gene duplications. Genome reorganization and horizontal gene transfer are sometimes associated with recombination at tRNA genes. Such genomic perturbations allow an increased evolutionary tempo. Thus RNase P may have coevolved with an increased rate of

horizontal gene transfer that facilitated the evolution of the genetic code.

## METHODS SUMMARY

Oligonucleotide synthesis and DNA sequencing was performed by the Keck Foundation Biotechnology Resource Laboratory at Yale University. *N. equitans* cells were obtained from K. O. Stetter and M. Thomm.

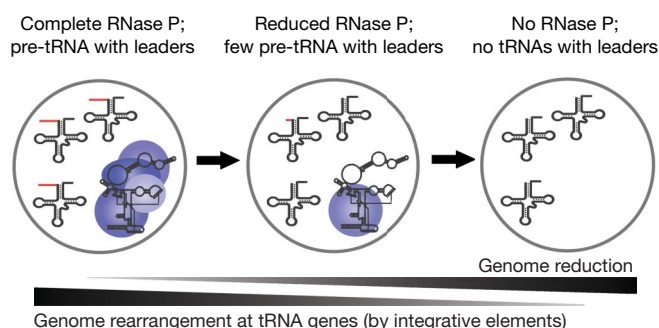
**Preparation and purification of RNA transcripts.** The *N. equitans* initiator pre-tRNA gene was cloned with the leader sequence 5'-GGUUAUAACUUACU-3' into a pUC19 vector that allowed for *in vitro* T7 RNA polymerase run-off transcription after plasmid cleavage with *Nsi*I. The tRNA was internally labelled in the presence of [ $\alpha$ -<sup>32</sup>P]ATP and purified as described<sup>29</sup>. RNase P RNA was synthesized *in vitro* from the plasmid pJA2' provided by C. Guerrier-Takada and S. Altman.

**Vaccinia virus capping enzyme assay.** Five micrograms of the total RNA of *M. kandleri*, *N. equitans* and *P. aerophilum* were incubated with 5 U of capping enzyme (Ambion) and 56 pmol [ $\alpha$ -<sup>32</sup>P]GTP (800 Ci mmol<sup>-1</sup>) for 1 h at 37 °C according to the manufacturer's instruction. The RNA was extracted with phenol:chloroform, ethanol precipitated in the presence of glycogen, washed with 70% ethanol, dried and run on a 12% acrylamide gel containing 8 M urea.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 November 2007; accepted 11 February 2008.

- Altman, S. *et al.* Catalysis by the RNA subunit of RNase P — a minireview. *Gene* **82**, 63–64 (1989).
- Evans, D., Marquez, S. M. & Pace, N. R. RNase P: interface of the RNA and protein worlds. *Trends Biochem. Sci.* **31**, 333–341 (2006).
- Vögeli, G., Grosjean, H. & Söll, D. A method for the isolation of specific tRNA precursors. *Proc. Natl Acad. Sci. USA* **72**, 4790–4794 (1975).
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
- Pannucci, J. A., Haas, E. S., Hall, T. A., Harris, J. K. & Brown, J. W. RNase P RNAs from some Archaea are catalytically active. *Proc. Natl Acad. Sci. USA* **96**, 7803–7808 (1999).
- Kikovska, E., Svard, S. G. & Kirsebom, L. A. Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proc. Natl Acad. Sci. USA* **104**, 2062–2067 (2007).
- Tsai, H. Y., Pulukkunat, D. K., Woznick, W. K. & Gopalan, V. Functional reconstitution and characterization of *Pyrococcus furiosus* RNase P. *Proc. Natl Acad. Sci. USA* **103**, 16147–16152 (2006).
- Li, Y. & Altman, S. In search of RNase P RNA from microbial genomes. *RNA* **10**, 1533–1540 (2004).
- Willkomm, D. K., Feltens, R. & Hartmann, R. K. tRNA maturation in *Aquifex aeolicus*. *Biochimie* **84**, 713–722 (2002).
- Klein, R. J., Misulovin, Z. & Eddy, S. R. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl Acad. Sci. USA* **99**, 7542–7547 (2002).
- Randau, L., Münch, R., Hohn, M. J., Jahn, D. & Söll, D. *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* **433**, 537–541 (2005).
- Palmer, J. R. & Daniels, C. J. *In vivo* definition of an archaeal promoter. *J. Bacteriol.* **177**, 1844–1849 (1995).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Hausner, W. & Thomm, M. in *Archaea, Evolution, Physiology and Molecular Biology* (eds Garrett, R. A. & Klenk, H. P.) 185–198 (Blackwell, Malden, Massachusetts, 2007).
- Kobayashi, T. *et al.* Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nature Struct. Biol.* **10**, 425–432 (2003).
- Farruggio, D., Chaudhuri, J., Maitra, U. & RajBhandary, U. L. The A1•U72 base pair conserved in eukaryotic initiator tRNAs is important specifically for binding to the eukaryotic translation initiation factor eIF2. *Mol. Cell. Biol.* **16**, 4248–4256 (1996).
- Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189–1232 (2002).
- Connolly, S. A., Rosen, A. E., Musier-Forsyth, K. & Francklyn, C. S. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. *Biochemistry* **43**, 962–969 (2004).
- Gupta, R. *Halobacterium volcanii* tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *J. Biol. Chem.* **259**, 9461–9471 (1984).
- Shuman, S., Surks, M., Furneaux, H. & Hurwitz, J. Purification and characterization of a GTP-pyrophosphate exchange activity from vaccinia virions. *J. Biol. Chem.* **255**, 11588–11598 (1980).



**Figure 4 | A scenario that allows the loss of RNase P.** The pre-tRNA leader sequences are indicated in red; RNase P proteins are indicated in blue.

21. Noren, C. J. *et al.* *In vitro* suppression of an amber mutation by a chemically aminoacylated transfer RNA prepared by runoff transcription. *Nucleic Acids Res.* **18**, 83–88 (1990).
22. Walker, S. C. & Engelke, D. R. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* **41**, 77–102 (2006).
23. Waters, E. *et al.* The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12984–12988 (2003).
24. Das, S., Paul, S., Bag, S. K. & Dutta, C. Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* **7**, 186 (2006).
25. She, Q., Shen, B. & Chen, L. Archaeal integrases and mechanisms of gene capture. *Biochem. Soc. Trans.* **32**, 222–226 (2004).
26. Altman, S., Kirsebom, L. & Talbot, S. Recent studies of ribonuclease P. *FASEB J.* **7**, 7–14 (1993).
27. Fitz-Gibbon, S. T. *et al.* Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc. Natl Acad. Sci. USA* **99**, 984–989 (2002).
28. Hartmann, E. & Hartmann, R. K. The enigma of ribonuclease P evolution. *Trends Genet.* **19**, 561–569 (2003).
29. Randau, L. *et al.* The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves non-canonical bulge–helix–bulge motifs of joined tRNA halves. *Proc. Natl Acad. Sci. USA* **102**, 17934–17939 (2005).
30. Sprinzl, M. & Vassilenko, K. S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **33**, D139–D140 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. O'Donoghue, J. Yuan and L. Sherrer for help and encouragement. This work was supported by grants from the National Institute of General Medical Sciences and the Department of Energy (D.S.) and the National Science Foundation (I.S.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.S. ([dieter.soll@yale.edu](mailto:dieter.soll@yale.edu)).



## METHODS

**Preparation of cell lysates and RNA cleavage assays.** *P. aerophilum* and *N. equitans* cells were each re-suspended in buffer containing 50 mM Tris·HCl (pH 7.5), 500 mM NaCl and 3 mM DTT, broken by sonication and subsequently centrifuged for 30 min at 30,000g. RNase P cleavage assays were performed with radioactively labelled tRNA transcript for 30 min at 37 °C in buffer containing 50 mM HEPES (pH 7.5), 20 mM MgCl<sub>2</sub> and 500 mM NH<sub>4</sub>Cl, either using *E. coli* RNase P RNA transcript or cell extracts. Cleavage assays with archaeal cell extracts were performed with variation of the incubation temperature from 37 °C to 80 °C, incubation time from 10 min to 3 h, and with variation of salt concentration from 10 to 100 mM MgCl<sub>2</sub> and from 100 to 1,500 mM NH<sub>4</sub>Cl.

**Preparation and purification of RNA transcripts.** The genes for *N. equitans* tRNA<sub>i</sub><sup>Met</sup>, tRNA<sub>i</sub><sup>Met</sup> ΔA-1, tRNA<sup>Tyr</sup> and tRNA<sup>Tyr</sup> ΔG-1 (including a hammer-head ribozyme to release a tRNA with a 5'-terminal C residue<sup>31</sup>) were cloned into the pUC19 vector, which allowed for *in vitro* T7 RNA polymerase run-off transcription after plasmid cleavage with *Nsi*I (tRNA<sub>i</sub><sup>Met</sup>) or *Bst*NI (tRNA<sup>Tyr</sup>). The tRNA transcripts were produced and purified as described<sup>29</sup>. 5' monophosphorylated tRNA was generated by incubating 10 μg tRNA transcript with 20 units of Antarctic Phosphatase (NEB) at 37 °C for 1 h followed by incubation with 20 units of T4 polynucleotide kinase (NEB) and 10 mM ATP at 37 °C for 1 h.

**Aminoacyl-tRNA synthetase aminoacylation assay.** The *N. equitans metS* gene (NEQ457) and *tyrs* gene (NEQ389) were amplified by PCR from genomic DNA. The *metS* gene was cloned into the *Xho*I/*Bln*I site of pET15b (Invitrogen) and the *tyrs* gene was cloned into the *Nde*I/*Xho*I site of pET20b (Invitrogen) to facilitate expression of the proteins in the *E. coli* BL21-Codon Plus (DE3)-RIL strain (Stratagene). Cultures were grown at 37 °C in Luria-Bertani medium supplemented with 100 μg ml<sup>-1</sup> ampicillin and 34 μg ml<sup>-1</sup> chloramphenicol. Expression of the recombinant proteins was induced by autoinduction<sup>32</sup>. Cells were re-suspended in buffer containing 50 mM HEPES (pH 7.5) and 500 mM NaCl, and broken by sonication. The fractions were flocculated at 80 °C for 30 min and centrifuged for 30 min at 20,000g. Aminoacylation was performed at 50 °C in a 50 μl reaction with 50 mM HEPES (pH 7.0), 50 mM KCl, 4 mM ATP, 15 mM MgCl<sub>2</sub> and 3 mM DTT. The methionylation assays contained 4 μM tRNA transcript, 15 μM [<sup>35</sup>S]methionine (more than 1,000 Ci mmol<sup>-1</sup>, diluted 1:100 with cold methionine) and 100 nM MetRS. The tyrosylation assays contained 2 μM tRNA transcript, 20 μM [<sup>14</sup>C]tyrosine (483 mCi mmol<sup>-1</sup>) and 100 nM TyrRS. Portions of 10 μl were removed at the time intervals indicated in Fig. 3, and radioactivity measured as described<sup>29</sup>.

**Reverse transcription and sequencing.** Purification of total small RNA from *P. aerophilum* and *N. equitans* cells, circularization, reverse transcription with Thermoscript (Invitrogen) reverse transcriptase at 70 °C, amplification and sequencing were performed as described<sup>11</sup>. The following primers were used for the amplification of tRNA<sup>Gly</sup>(TCC): forward 5'-GCGCCCGTAGTCTAG-TGGTAGGATG-3', reverse 5'-TGCGCCCGCCGGGATTCGAACCCGG-3'; tRNA<sup>Tyr</sup>: forward 5'-GCCGGGCGTAGCTCAGCGGCAGAG-3', reverse 5'-TCCGGGCGGGGGGATTCGAACCC-3'; and tRNA<sup>Thr</sup>(TGT): forward 5'-GCCCCGGTAGCTCAGCGGCAGAGCG-3', reverse 5'-AGCCCCGGGCGGG-ATTCGAACCC-3'. Circularization of total tRNA of *N. equitans* and *P. aerophilum* was performed with Thermophage ssDNA ligase (Prokaria) at 65 °C according to the manufacturer's directions. The following primers were used for the amplification of circularized tRNA<sub>i</sub><sup>Met</sup>: forward 5'-TCATAACCCCCAGGT-CCCCGGTTCAAATCCG-3', reverse 5'-TATGAGCCCCCGGGCACTCCA-GGCTGCC-3'; tRNA<sup>Tyr</sup>: forward 5'-TGTAGACCGGCAGGTCCGGGGTTCG-AATCCC-3', reverse 5'-CTACAGCCGGCCGCTCTGCCGCTGAGC-3' and tRNA<sup>His</sup>: forward 5'-TGTGGGACCCGGAGGTCCCGGGTTCGAATCCC-3', reverse 5'-CCACAGCCCGGCGCTCTGCCACTAAGC-3'.

31. Fechter, P., Rudinger, J., Giegé, R. & Théobald-Dietrich, A. Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. *FEBS Lett.* **436**, 99–103 (1998).

32. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

## LETTERS

# Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS

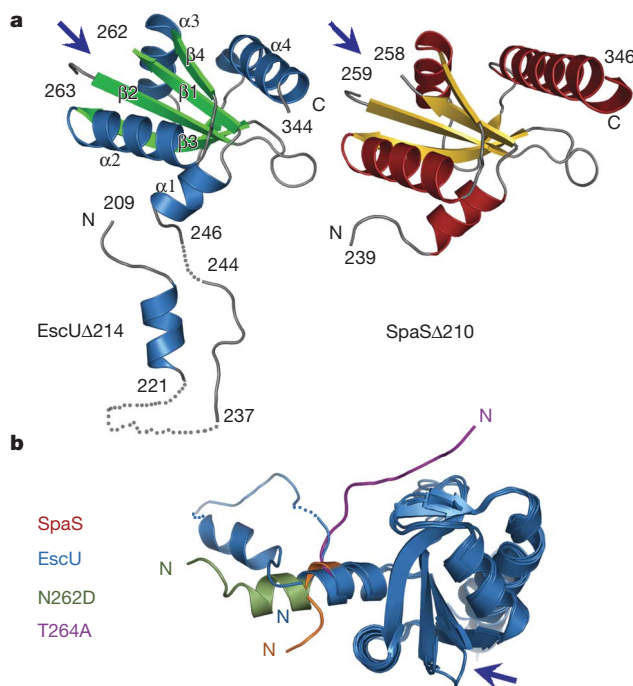
Raz Zarivach<sup>1</sup>, Wanyin Deng<sup>2</sup>, Marija Vuckovic<sup>1</sup>, Heather B. Felise<sup>3</sup>, Hai V. Nguyen<sup>3</sup>, Samuel I. Miller<sup>3</sup>, B. Brett Finlay<sup>2</sup> & Natalie C. J. Strynadka<sup>1</sup>

During infection by Gram-negative pathogenic bacteria, the type III secretion system (T3SS) is assembled to allow for the direct transmission of bacterial virulence effectors into the host cell<sup>1</sup>. The T3SS system is characterized by a series of prominent multi-component rings in the inner and outer bacterial membranes, as well as a translocation pore in the host cell membrane. These are all connected by a series of polymerized tubes that act as the direct conduit for the T3SS proteins to pass through to the host cell. During assembly of the T3SS, as well as the evolutionarily related flagellar apparatus<sup>2</sup>, a post-translational cleavage event within the inner membrane proteins EscU/FlhB is required to promote a secretion-competent state. These proteins have long been proposed to act as a part of a molecular switch, which would regulate the appropriate chronological secretion of the various T3SS apparatus components during assembly and subsequently the transported virulence effectors. Here we show that a surface type II  $\beta$ -turn in the *Escherichia coli* protein EscU undergoes auto-cleavage by a mechanism involving cyclization of a strictly conserved asparagine residue<sup>3</sup>. Structural and *in vivo* analysis of point and deletion mutations illustrates the subtle conformational effects of auto-cleavage in modulating the molecular features of a highly conserved surface region of EscU, a potential point of interaction with other T3SS components at the inner membrane. In addition, this work provides new structural insight into the distinct conformational requirements for a large class of self-cleaving reactions involving asparagine cyclization.

The EscU family of proteins is an essential component of the inner membrane ring of T3SS systems. It is thought to be composed of approximately half a dozen highly conserved proteins (EscJNRSTUV in enteropathogenic *E. coli*). Earlier primary sequence analysis of the EscU and orthologous FlhB (flagellar) family of proteins predicted two major domains: an approximate 200-amino-acid amino (N)-terminal domain (NTD) that is involved in the association to the T3SS inner membrane basal ring; and an approximate 100-amino-acid carboxy (C)-terminal domain (CTD), which is known to undergo auto-cleavage and to mediate a switch in the secretion substrate profile of the T3SS<sup>4</sup> (Supplementary Fig. 1). The two domains are connected by a proposed 30-amino-acid flexible linker, which is highly conserved (more than 80%) within the EscU/SpaS and FlhB family of proteins (Supplementary Fig. 2). In this study, constructs encompassing the predicted linker and CTD switch domain from the T3SS EscU protein of enteropathogenic *E. coli* (EPEC), residues 215–345 termed EscUA $\Delta$ 214, and the SpaS protein from *Salmonella typhimurium*, residues 211–356 termed SpaSA $\Delta$ 210, were purified and crystallized. Several crystal forms from the two species were obtained with diffraction up to a resolution of

1.2 Å (Fig. 1a; for simplicity, EscU numbering is used throughout the text).

The crystallographic data collectively show that the CTD domains are highly conserved in structure (despite the modest approximate 30% sequence identity of the region) (Supplementary Fig. 3). The root mean square deviations between all CTD structures of EscU from seven different space groups are approximately 0.6 Å for the 93 backbone atoms. A root mean square deviation of 1.0 Å was observed for the 94 common backbone atoms of EPEC EscU and *Salmonella* SpaS, indicating that the CTD has a compact stable fold including highly similar conformations of the several surface loops (Fig. 1). The only significant deviation between the CTDs of EscU and SpaS is the elongated C-terminal  $\alpha$ -helix, which is extended by a



**Figure 1 | Structure of the C-terminal domains of EscU and SpaS.** **a**, The native cleaved CTD of EscU and SpaS with a blue arrow pointing to the auto-cleavage site. CTD is a novel  $\alpha/\beta$ -fold with a mixed parallel and anti-parallel five-stranded twisted  $\beta$ -sheet (topology  $\beta_4, \beta_1, \beta_2, \beta_3, \beta_5$ ) flanked by two helices on each side ( $\alpha_1, \alpha_2, \alpha_3$  and  $\alpha_4$ ). **b**, Superposition of the CTD reveals a different fold for the N-terminal linker between EscU, EscU mutants and SpaS, as well as a longer C-terminal helix for SpaS.

<sup>1</sup>Department of Biochemistry and Molecular Biology, and the Center for Blood Research, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada. <sup>2</sup>Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada. <sup>3</sup>Department of Microbiology and Medicine, HSB K-140, Box 357710, Seattle, Washington 98195, USA.

15-residue sequence insertion in SpaS. As supported by light scattering analysis and mass spectrometry (data not shown), the EscU and SpaS CTD constructs are strongly monomeric both in solution and in our several crystal lattices.

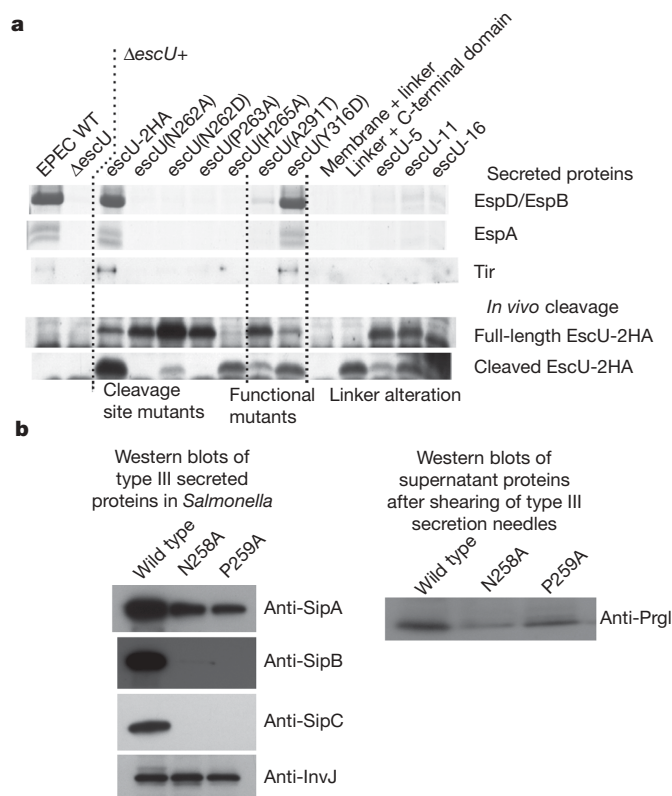
EscU contains a cleavage site conserved within all T3SS and flagellar orthologues<sup>5</sup>. The site of auto-cleavage was verified structurally here and is localized to an exposed region between strands  $\beta$ 1 and  $\beta$ 2 containing a highly conserved sequence quartet observed in all T3SS and flagellar EscU/FlhB variants (NPTH defined by Asn262, Pro263, Thr264, His265 in EscU). We also show that purification of the CTD of EscU in the absence of the membrane anchoring NTD did not obliterate the self-cleavage process as observed directly in the high-resolution structures (Figs 1 and 3a). In our structures, self-cleavage uniformly occurred at the predicted Asn262–Pro263 peptide, liberating the termini of the newly created peptides locally at the protein surface, but clearly not interfering with the overall integrity of the protein fold (Supplementary Fig. 4).

To address the mechanism of cleavage further, we analysed six point mutations at the auto-cleavage site in EscU (N262A, N262D in two crystal forms (that is, N262D and N262Di), P263A, T264A, H265A and R313T) (Supplementary Figs 4 and 9). By overlapping the structures of our native and mutant forms of EscU, we were able to elucidate the molecular details of a self-cleavage intein-like mechanism involving asparagine cyclization (Supplementary Fig. 12)<sup>6–8</sup>. The mechanism initiates during the post-translational folding of EscU, including creation of the type II  $\beta$ -turn (T2 $\beta$ ) at residues 262–265. In this initial uncleaved form, the carbonyl oxygen of the

strictly conserved Asn262 (notably localized in position one of the T2 $\beta$  in our structures and not at position two as previously hypothesized)<sup>9,10</sup> is hydrogen bonded to the backbone amide of His265 and Ile266. The T2 $\beta$  unique to the uncleaved form enforces a 90° angle between the Asn262–Pro263 amide to Pro263–Thr264 amide (Asn262,  $\psi = 112$ ; Pro263,  $\phi = -47$ ;  $\psi = 133$ )<sup>11</sup>, creating a distance of about 2.8 Å between the carbonyl oxygen of Asn262 and the carbonyl carbon of Pro263, and an  $n \rightarrow \pi^*$  interaction which would promote electron withdrawal from the Asn262 carbonyl oxygen (Fig. 3 and Supplementary Fig. 5)<sup>12</sup>. Supporting the need for this strained conformation in catalysis is the observation that the cleavage-impaired mutant P263A (it promotes very low levels of cleavage only at conditions of extreme pH<sup>13</sup>) adopts a type I rather than T2 $\beta$  turn owing to the loss of the requisite main-chain conformation imposed by the conserved proline residue (Asn262,  $\psi = 104$ ; Pro263,  $\phi = 51$ ;  $\psi = -128$ ). The type I conformation does not promote the favourable hydrogen bonding and  $n \rightarrow \pi^*$  interactions that we propose are essential to efficient catalysis. As the reaction initiates, the side-chain N $\delta$  of Asn262 is exposed on the protein surface and enables water molecules to absorb a proton to promote the attack of the N $\delta$  lone pair on the partly positively charged carbonyl carbon, creating a tetrahedral intermediate. The negative charge that develops on the Asn262 carbonyl oxygen in this intermediate form is stabilized by hydrogen bonds to the backbone amides of His265/Ile266 and by electron withdrawal by the  $n \rightarrow \pi^*$  interaction. Collapse of the tetrahedral intermediate and subsequent cleavage are enabled by absorbing back the proton from water to the leaving amino group at Pro263, creating a new N terminus at Pro263 and a C-terminal succinimide intermediate at Asn262. The succinimide intermediate (Supplementary Fig. 6; P263Asuc in Supplementary Table 1) can subsequently be hydrolysed to create the new C-terminal negatively charged Asn262 (Fig. 3b). We can deduce, based on our non-cleaved mutant structures, that the asparagine driven auto-cleavage mechanism relies on the *trans* orientation between the Asn carbonyl and the following peptide carbonyl, and stabilization of the Asn carbonyl by hydrogen bonding to an electro-positive acceptor.

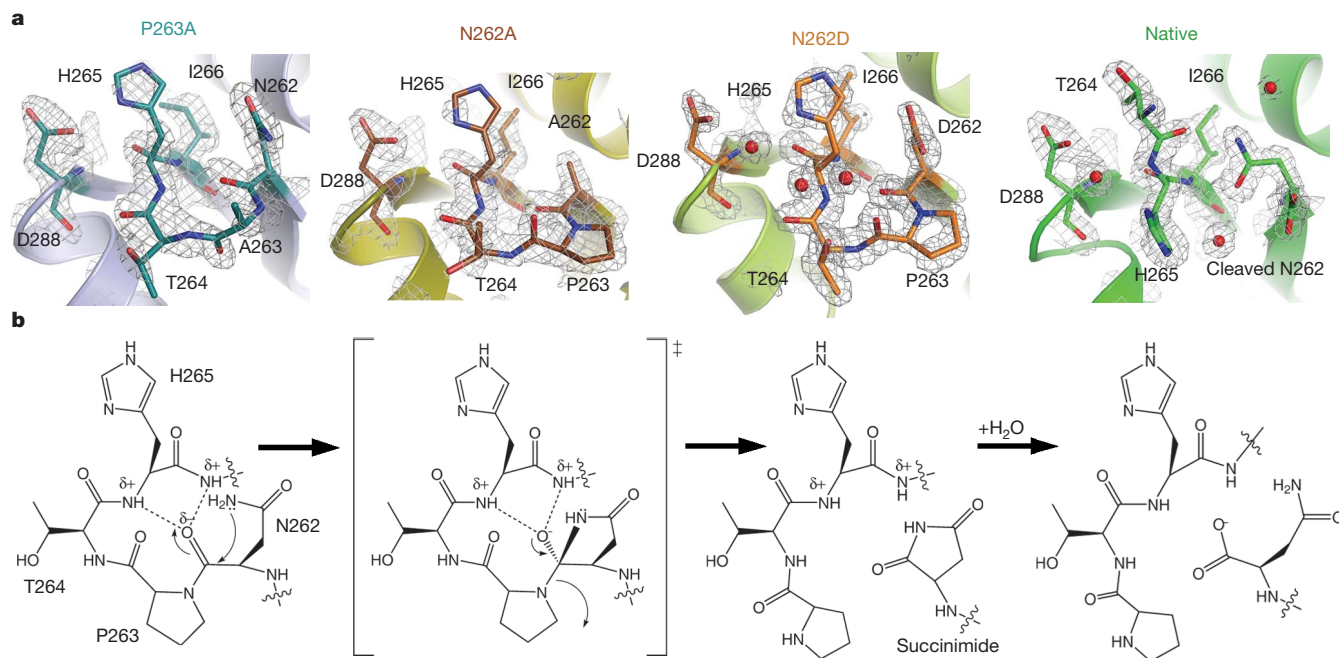
To assess the importance of the conserved residues (Supplementary Fig. 3) to the T3SS function, we have tested EscU and SpaS non-cleavable mutants for secretion of T3SS substrates (Fig. 2). In agreement with earlier observations in *Yersinia*<sup>4</sup>, in EPEC and *Salmonella* we observe full arrest of secretion of the translocon and filament components (Fig. 2a). However, in contrast to the *Yersinia* study<sup>4</sup>, the SpaS mutants did not alter the secretion level of the molecular ruler (InvJ), but altered the secretion level of the needle component (PrgI) (Fig. 2b). Thus, our data suggest that EscU/SpaS act in the recognition of the needle components and as part of a switch between the secretion of the needle and the T3SS translocon (SipB/C in *Salmonella*) or filament (EspA in *E. coli*) components.

In addition to electrostatic features (Supplementary Fig. 7), additional cleavage-induced conformational features likely govern the cleaved, T3SS competent state. The breaking of the amide bond between Asn262 to Pro263 leads to a conformational change of the newly created N- and C-terminal ends from their positions within the T2 $\beta$  pre-cleavage to an extended form with completely distinct side-chain orientations post-cleavage (Fig. 3a). The newly introduced N- and C-terminal charges also contribute to the altered electrostatic profile of this highly conserved surface. The dispositions of critical side chains in the loop are also significant. For example, we have shown that an H265A mutant is fully cleaved but is unable to promote secretion (Fig. 2a), implying a direct role for the imidazole side chain in the secretion process. In the cleaved form the side chain of Asn262 is hydrogen bonded to the backbone of the now flipped strand containing Thr264 to Ile266, creating a more extended  $\beta$ -chain and effectively re-orienting the critical His265 by about 180°, transplanting it from one face of EscU to another (Fig. 3a). Post-cleavage, the His265 imidazole side chain sits at the centre of a



**Figure 2 | Type III secretion in EPEC and *S. typhimurium*.** **a**, EPEC, *escU* mutants and complementation of type III secretion. Secretion of effector (Tir) and translocator proteins (EspB/D) checked under complementation of *escU* by WT and mutants. The cleavage state of EscU *in vivo* is presented for all EscU components. **b**, *spaS* WT, mutants and type III secretion. Secretions of effector (SipA), molecular ruler (InvJ), needle (PrgI) and translocator proteins (SipB/C) were analysed in WT and *spaS* mutants. The N258A and P259A substitutions in SpaS correspond to N262A and P263A respectively in EscU. Full-length blots/gels are presented in Supplementary Figure 1.





**Figure 3 | Auto-cleaving mechanism of EscU.** **a**, Non-cleaved type I  $\beta$ -bend of P263A (left), the non-cleaved type II beta bend of N262A (middle left), the non-cleaved type II beta bend of N262Di (middle right) and the native cleaved loop with a flipped His 265 (right). Note the identical conformations

of the N, C, C $\alpha$  and C $\beta$  at position 262 in all uncleaved forms. All maps are sigma-A weighted  $2F_o - F_c$  electron density ( $1.5\sigma$ ); water molecules are represented by the red spheres. **b**, Detailed mechanism for the asparagine cyclization in EscU.

newly created surface formed by (in addition to the N- and C-terminal charges at positions 262 and 263) the conserved Arg313, Leu292 and positively charged Lys261 (EscU) or Arg295 (SpaS). Collectively the cleavage creates a localized but electrostatically and conformationally unique surface that we predict is essential for the interaction with other T3SS components.

EscU and orthologues in other T3SSs, as well as the flagellar system, have been implicated in acting as part of a molecular 'switch' that regulates the chronological secretion of apparatus components and virulence effectors (detailed in Supplementary Fig. 10). Although our structural and mutational data of the conserved cleavage products show them to be attractive interaction sites for the other T3SS components globally constituting the inner membrane ring, and our thermal denaturation data directly indicate that cleavage is a critical event for stabilization of these inner membrane proteins (Supplementary Fig. 11), it is clear that the post-translational auto-cleavage, which occurs immediately after translation and folding in the cytoplasm, although essential to secretion, is not the switching event per se.

Our structural and biochemical data point to the conserved linker region of EscU orthologues for playing a potential role in the switching events. The proposed linker region (EscU 215–245 and SpaS 210–241) connects the N-terminal membrane-anchored domain with the globular CTD. Although earlier sequence analysis predicted a helical linker<sup>14</sup>, all our structures show a large part of the linker is unstructured. Overlay of the SpaS and EscU structures (Fig. 1b) indicates conformational differences and potential flexibility in the linker that centre on a hinge-like region between the end of the linker and the initial helix of the CTD (Fig. 1b). This hinge provides a potential point of conformational freedom for the CTD from the membrane anchored N-terminal domain of EscU and its associated T3SS partners. One can envisage such a hinging motion could allow for the binding and dissociation of the EscU CTD from the T3SS inner membrane base. This would effectively promote its role as a molecular switch by appropriately changing the entrance shape, size and electrostatic nature of the T3SS pore as downstream T3SS

components and effectors are bound and secreted in the specific chronology that has been previously observed<sup>4</sup>.

To analyse the importance of the observed linker in T3SS function, we used our structure to design three deletion mutations (EscU(–5) 232–236, EscU(–11) 234–245, EscU(–16) 230–245) and two point mutations (G229P and G235P) in the flexible hinge region of EscU. These mutants were screened *in vivo* for secretion of the filament component EspA, the translocon components (EspB/EspD) and for the effector Tir in *AescU* and in wild-type (WT) backgrounds (Fig. 2, and Supplementary Figs 8 and 13). In all cases, alteration of the linker length abolishes the secretion of the translocon components and effector in the *AescU* knockout and decreased their secretion in the WT strain (a finding also consistent with previous mutagenesis studies in the flagellar FlhB family<sup>14</sup>). Furthermore, G229P and G235P show dramatic alteration of T3SS secretion (Supplementary Fig. 13). Given our observations that these EscU mutants are expressed and folded properly *in vivo* (owing to their ability to perform auto-cleavage), the observed *in vivo* loss of function of the linker deletions and point mutations supports the proposed requirement for the appropriate juxtaposition of the auto-cleavable CTD relative to the membrane-bound NTD of EscU for the binding, localization and secretion of the T3SS assembly and effector molecules.

## METHODS SUMMARY

Protein purification, site-directed mutagenesis and secretion assays were performed as previously published<sup>15</sup> with small alterations (see Methods). Crystallization and crystallographic statistics are in Supplementary Tables 1 and 2.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 August 2007; accepted 12 February 2008.

1. Hueck, C. J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
2. Macnab, R. M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).

3. Ferris, H. U. & Minamino, T. Flipping the switch: bringing order to flagellar assembly. *Trends Microbiol.* **14**, 519–526 (2006).
4. Sorg, I. *et al.* YscU recognizes translocators as export substrates of the *Yersinia* injectisome. *EMBO J.* **26**, 3015–3024 (2007).
5. Lavander, M. *et al.* Proteolytic cleavage of the FlhB homologue YscU of *Yersinia pseudotuberculosis* is essential for bacterial survival but not for type III secretion. *J. Bacteriol.* **184**, 4500–4509 (2002).
6. Ding, Y. *et al.* Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* **278**, 39133–39142 (2003).
7. Paulus, H. Protein splicing and related forms of protein autoprocesing. *Annu. Rev. Biochem.* **69**, 447–496 (2000).
8. Shemella, P. *et al.* Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations. *Biophys. J.* **92**, 847–853 (2007).
9. Geiger, T. & Clarke, S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* **262**, 785–794 (1987).
10. Stephenson, R. C. & Clarke, S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J. Biol. Chem.* **264**, 6164–6170 (1989).
11. Clarke, S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. *Int. J. Pept. Protein Res.* **30**, 808–821 (1987).
12. Hodges, J. A. & Raines, R. T. Energetics of an  $n \rightarrow \pi^*$  interaction that impacts protein structure. *Org. Lett.* **8**, 4695–4697 (2006).
13. Ferris, H. U. *et al.* FlhB regulates ordered export of flagellar components via autocleavage mechanism. *J. Biol. Chem.* **280**, 41236–41242 (2005).
14. Fraser, G. M. *et al.* Substrate specificity of type III flagellar protein export in *Salmonella* is controlled by subdomain interactions in FlhB. *Mol. Microbiol.* **48**, 1043–1057 (2003).
15. Zarivach, R., Vuckovic, M., Deng, W., Finlay, B. B. & Strynadka, N. C. Structural analysis of a prototypical ATPase from the type III secretion system. *Nature Struct. Mol. Biol.* **14**, 131–137 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T.S., M.C. and P.I.L. for discussions; the staff at the Advanced Light Source beamline 8.2.2 and SSRL beamline 11-1 for data collection time and assistance; and E. Galyov at the Institute for Animal Health for the SipA, SipB and SipC antibodies. This work was supported by an Izaak Walton Killam Research post-doctoral fellowship, a Michael Smith Foundation for Health Research (MSFHR) post-doctoral fellowship and a Canadian Institutes of Health Research (CIHR) post-doctoral fellowship (all to R.Z.). N.C.J.S. and B.B.F. thank the Howard Hughes International Scholar program and the CIHR for funding. N.C.J.S. also thanks the MSFHR and the Canada Foundation of Innovation for infrastructure funding support. N.C.J.S. is also an MSFHR Senior Scholar and CIHR Investigator. S.M. acknowledges grants from the National Institutes of Allergy and Infectious Diseases (NIAID), U54 AI057141 and 5RO1 AI030479.

**Author Contributions** R.Z. and M.V. cloned and purified EscU and SpaS. R.Z. crystallized EscU and SpaS. R.Z. solved the structures. H.B.F. isolated the chromosomal SpaS mutants. W.D., H.B.F. and H.V.N. performed the biochemical experiments in Fig. 2. B.B.F. and S.I.M. provided resources for the experiments shown in Fig. 2. N.C.J.S. provided resources for all data other than those in Fig. 2. All authors discussed the results and commented on the manuscript.

**Author Information** The atomic coordinates of all the structures have been deposited in the Protein Data Bank with the accession codes 3BZL, 3BZO, 3BZP, 3BZR, 3BZS, 3BZT, 3C03, 3BZV, 3BZX, 3BZY, 3BZZ, 3C00 and 3C01. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to N.C.J.S. ([natalie@byron.biochem.ubc.ca](mailto:natalie@byron.biochem.ubc.ca)).

## METHODS

**Protein expression and purification.** For purification of His-tagged SpaΔ210, EscUΔ214 and all mutants, the expression construct pEscU/pSpaS was generated by PCR-amplifying the *escU/spaS* open reading frame from EPEC E2348/69 and *S. typhimurium* genomic DNA, and cloning it into the *NheI/BamHI* and *NheI/HindIII* sites of the pET-28(a) expression vector (Novagen). *E. coli* BL21 (DE3) transformed with pEscU/pSpaS was grown to mid-exponential phase at 37 °C in Luria-Bertani broth containing kanamycin and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside. Cells were harvested after incubation for 16 h at 20 °C, re-suspended in buffer (20 mM Tris, 500 mM NaCl, pH 8.0), lysed using a French press, and centrifuged at approximately 60,000g for 35 min. His-tagged EscU/SpaS was purified from the soluble fraction using nickel-chelating Sepharose (Amersham). Protein was eluted using the buffer (20 mM Tris, 40 mM NaCl, 300 mM imidazole, pH 8.0) and the tag cleaved at 4 °C overnight with 1:1,000 thrombin (Sigma), leaving six extra residues at the N terminus (Gly-Ser-His-Met-Ala-Ser). Cleaved product was purified by Mono-Q anion-exchange (Pharmacia) using a linear gradient of NaCl in 20 mM Tris-HCl buffer, pH 8.0.

**Site-directed mutagenesis by PCR.** EscU mutants were generated using QuikChange site-directed mutagenesis (coding and antisense primers containing a single mutagenic site were used for PCR amplification).

**Crystallization and structure determination.** Purified EscUΔ214/SpaΔ210 (at 15–20 mg ml<sup>-1</sup>) crystallized into various forms in a variety of conditions. Crystals were grown at 21 °C using sitting-drop vapour diffusion or under oil microbatch by mixing 0.5 μl of protein with the same volume of reservoir solution (Supplementary Table 2). Crystal screening and data collection were done at Beamline 8.2.2 of the Advanced Light Source, 11-1 of the Stanford Synchrotron Radiation Laboratory, and Cu Kα at home source. Data were reduced and scaled using the HKL2000<sup>16</sup> suite. For phasing, EscUΔ214 crystals were soaked into 2 M NaI 2 M Na formate for 2 min and flash frozen in liquid nitrogen. Positions of all the seven iodide sites were found and refined by ShelX-C<sup>2</sup>. Phase calculation and solvent flattening were done using ShelX-DE<sup>17</sup>, which resulted in an interpretable map. The model was built using Coot<sup>18</sup>, and refinement was done using Refmac after excluding 5% of the data for the R-free calculation<sup>19</sup>. SpaΔ210 and EscUΔ214 mutants were solved by molecular replacement using Phaser<sup>20</sup>. Some residues were not observed in the crystal (Supplementary Table 1). Structural figures were prepared with PyMOL<sup>21</sup> and electrostatic calculations were done with the APBS plugin<sup>22</sup>.

**Generation of the ΔescU mutant in EPEC.** The *sacB* gene-based allelic exchange method<sup>23</sup> was used to generate an *escU* in-frame deletion mutant in the streptomycin-resistant derivative (Sm<sup>r</sup>) of EPEC O127:H6 strain E2348/69 using the suicide vector pRE112. To make a deletion mutant of *escU* in EPEC, PCR was used to generate two fragments (1.1 and 1.3 kilobases, respectively) using primer pairs EPescU-1 (*KpnI*) (5'-GGGTACCTTATGTGTGCAACGTTCTGG-3') and DEPescU-R (*NheI*) (5'-CGCTAGCTTCACTTTTGTACATCGCC-3') as well as DEPescU-F (*NheI*) (5'-CGCTAGCTTGATTGCGATAGAC-3') and EPescU-2 (*SacI*) (5'-GGAGCTCCTTCGGCAATATCATTGCGAG-3'). The PCR products were cloned into pCR2.1-TOPO (Invitrogen) and verified by DNA sequencing. After digestion with *KpnI/NheI* and *NheI/SacI*, respectively, the two fragments were gel-purified and cloned into pRE112 digested with *KpnI/SacI* in a three-way ligation. The resulting plasmid pRE-ΔEPescU contained 1–2 kilobases of flanking regions on both sides of *escU* and the *escU* gene with an internal in-frame deletion from nucleotides 79 to 1007 (about 89% of the coding region). An *NheI* site was introduced into the deletion site. Plasmid pRE-ΔEPescU was transformed into *E. coli* SM10λpir by electroporation, and introduced into EPEC strain E2348/69 Sm<sup>r</sup> by conjugation. After sucrose selection, EPEC colonies resistant to sucrose and sensitive to chloramphenicol were screened for deletion of *escU* by PCR using primers EPescU-1 and EPescU-2. The EPEC *escU* mutants were further verified by multiple PCR reactions.

**Expression plasmid of EPEC *escU* for complementation.** Primers EPescU-HAF (*SacI*) (5'-CGAGCTCACGGCAAATATTCATTCTGAC-3') and EPescU-HAR (*XhoI*) (5'-CCTCGAGATAATCAAGGTCTATCGCAATAC-3') were used to amplify by PCR a fragment containing the coding region as well as a 55-bp upstream region of *escU*. The PCR product was cloned into pCR2.1-TOPO, verified by DNA sequencing, and subcloned into the vector pTOPO-2HA<sup>24</sup> digested with *SacI/XhoI* to generate plasmid pEPescU-2HA. This plasmid expresses EscU with a double haemagglutinin tag (2HA) at the C terminus under the control of the *Plac* promoter on the vector, and can complement the EPEC ΔescU mutant. The expression of EscU conferred by the plasmid was followed by western blotting using the haemagglutinin monoclonal antibody (Covance).

**Type III secretion assay for EPEC.** EPEC strains were grown in Luria-Bertani and then subcultured in Dulbecco's Modified Eagle Media to induce type III secretion as previously described<sup>15</sup>.

**Strain construction for *S. typhimurium*.** The following genetic manipulations were performed using the λ-RED system<sup>25</sup>. The tetracycline resistance gene *tetA* and its transcriptional repressor *tetR* were introduced into the *spaS* gene in WT *S. typhimurium* after amplification with the following primers: HF315, 5'-GGTGAATCTGATATTGAAAACACGCGCTGATTGTTCCTTAAGACCCCACTTTCACA; HF316, 5'-TCGGCATCAATTCGGGTTTAAAAATAAATCCCGATCGTAATCTAAGCACTTGTCTCTCTG-3'. The tetracycline resistance genes replaced the coding sequence for amino acids 259–261 of *spaS*. Allelic exchange and selection for tetracycline sensitivity<sup>26</sup> was used to replace the tetracycline resistance with the N258A and P259A site-directed mutants amplified with the following primers: HF317, 5'-GATATTGAAAACACGCGCTGATTGTTCGCGCCCCACGCATATTACGATCGGGATT-3'; HF318, 5'-ATAAATCCCGATCGTAATATGCGTGGGGCGGCAACAATCAGGCGTGAGTTTTCATATC-3' for N258A; HF319, 5'-GATATTGAAAACACGCGCTGATTGTTCGCAACGCCACGCATATTACGATCGGGATT-3'; HF320, 5'-ATAAATCCCGATCGTAATATGCGTGGCGTTGGCAACAATCAGGCGTGAGTTTTCATATC-3' for N259A.

**Western blots for *S. typhimurium*.** Secreted proteins were prepared as previously described for *S. typhimurium*<sup>27</sup> and SDS-polyacrylamide gel electrophoresis and western blot techniques were performed as described<sup>28</sup>.

**Needle shearing for *S. typhimurium*.** Overnight cultures of WT *S. typhimurium* and the *spaS* mutants were grown in Luria-Bertani medium at 37 °C. Cultures were passed six times through a hypodermic needle (23Gx1, 0.6 mm × 25 mm; B. Braun) to release surface needle proteins. The cultures were centrifuged at 1,800g for 15 min, and the supernatants were precipitated with 10% trichloroacetic acid for 20 min at 4 °C. The samples were centrifuged at 1,800g for 20 min at 4 °C, and the pellets were re-suspended in 50 mM Tris, pH 8.0, and sample buffer. Equal amounts of supernatant proteins were loaded for each strain, and SDS-polyacrylamide gel electrophoresis and western blot techniques were performed as described<sup>28</sup>.

**Circular dichroism spectroscopy.** The circular dichroism spectra were recorded with a Jasco spectropolarimeter (Model J-810) equipped with a Pelletier device. All spectra are averages of three scans, quartz cells of optical path length 0.2 cm and concentrations of 2.5 and 5 μM. EscUΔ214 and mutants in buffer (0.1 mM Tris, pH 8.0) were used. For thermal denaturation experiments, melting curves were determined by monitoring the changes in dichroic density at 222 nm as a function of temperature in the range 25–90 °C and at a heating rate of 1 °C min<sup>-1</sup>. The thermodynamic parameters associated with the temperature-induced denaturation were obtained by nonlinear, least-squares analysis of the temperature dependence of circular dichroism. A two-state denaturation process was assumed during curve-fitting analyses.

- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).
- Schneider, T. & Sheldrick, G. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
- PyMOL. v.1 (DeLano Scientific, Palo Alto, California, 2002).
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
- Edwards, R. A., Keller, L. H. & Schifferli, D. M. Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene* **207**, 149–157 (1998).
- Deng, W. et al. Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc. Natl Acad. Sci. USA* **101**, 3597–3602 (2004).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
- Maloy, S. & Nunn, W. Selection for loss of tetracycline resistance by *Escherichia coli*. *J. Bacteriol.* **145**, 1110–1112 (1981).
- Kimbrough, T. G. & Miller, S. I. Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation. *Proc. Natl Acad. Sci. USA* **97**, 11008–11013 (2000).
- Pegues, D. A., Hantman, M. J., Behlau, I. & Miller, S. I. PhoP/PhoQ transcriptional repression of *Salmonella typhimurium* invasion genes: evidence for a role in protein secretion. *Mol. Microbiol.* **17**, 169–181 (1995).



## CORRIGENDUM

[doi:10.1038/nature06946](https://doi.org/10.1038/nature06946)**Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila***

Cathy Haag-Liautard, Mark Dorris, Xulio Maside, Steven Macaskill, Daniel L. Halligan, David Houle<sup>1</sup>, Brian Charlesworth & Peter D. Keightley

<sup>1</sup>Department of Biological Science, Florida State University, Tallahassee, Florida 32306-1100, USA.

*Nature* 445, 82–85 (2007)

---

In this Letter, David Houle was omitted from the author list. David Houle was responsible for producing the Florida mutation accumulation lines that were analysed in the experiment.

# naturejobs

**JOBS OF  
THE WEEK**

**T**he US National Postdoctoral Association is not pleased with salary trends for postdocs funded by the National Institutes of Health (NIH) — and, in an unprecedented move for the association, is taking its case to Congress.

In 2000, the National Academy of Sciences recommended that NIH postdoc salaries, through the NIH's postdoc support mechanism known as the National Research Service Award (NRSA), should rise by 10–12% per year until they reach a target salary of US\$45,000 for entry-level postdocs. The NIH agreed with this recommendation and added that once postdoc stipends had reached the target, they should continue to rise to keep pace with inflation and cost-of-living adjustments.

But the promised rises have not materialized. Salaries have risen by as much as 10% in some years since fiscal year 2002, but remained the same in 2007 and 2008. For fiscal year 2009, the US president has requested a 1% rise. The first-year salary for NIH postdocs is currently just \$36,996. The National Postdoctoral Association went to the NIH, but the agency referred it to Congress. Now, the association is encouraging its members to write letters to their Representatives and Senators. The association has a case, especially as so many institutions base their postdoc salaries on the NRSA's stipends.

But increased salaries are not necessarily the best way to improve job satisfaction. Results from a survey carried out by the science society Sigma Xi suggest that workplace satisfaction is linked only weakly to compensation levels. This, according to the society, is in line with the view that postdocs are driven more by future employment prospects than by their current salary. More important for job satisfaction are factors such as training opportunities and the amount of forward planning that new postdocs engage in with their advisers. Postdocs deserve good pay. But perhaps the more rewarding — and possibly tougher — battles would be for better training, better advising and better future job opportunities.

**Gene Russo is editor of *Naturejobs*.**

## CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**

Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)  
**Southwest UK/RoW:** Nils Moeller (4953)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)  
**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)  
**North Germany/The Netherlands:**  
Reya Silao (4970)  
**South Germany/Austria:**  
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:** Dennis Chu

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718

Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**  
Vikas Chawla (+91 1242881057)  
e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe (+81 3 3267 8765)  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)  
**Business Development Manager, Greater China/Singapore:**  
Gloria To (+852 2811 7191)  
e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# MOVERS

**Eaton E. Lattman, chief executive and executive director of the Hauptman-Woodward Medical Research Institute, Buffalo, New York**



**2004–present:** Dean of Research and Graduate Education, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, Maryland

**1996–2004:** Chair, Department of Biophysics, Krieger School of Arts and Sciences, Johns Hopkins University

After spending almost his entire academic career at Johns Hopkins University as a student, researcher and administrator, Eaton Lattman has been lured away. The job, directing the Hauptman-Woodward Medical Research Institute, is a perfect fit, he says, because the institute's research profile mirrors his work in structural biology.

But an ambitious crystallography project, attempted at Johns Hopkins school of medicine in the 1980s, almost derailed Lattman's academic career. Ignoring advice that it would prove too difficult, he ambitiously tried to crystallize a virus called polyoma. He learned to grow mammalian cells and culture viruses, but couldn't grow crystals of sufficient quality for crystallography. His tenure was in jeopardy. Needing results quickly, Lattman began studying mutations that altered the folding of an already crystallized small protein: staphylococcal nuclease. After publishing a number of key papers in a short time, he earned tenure, and began studying the structure and function of proteins and RNA.

Lattman has long appreciated crystallography's challenges. After earning his BSc in chemistry and physics from Harvard College in 1962, he chose a PhD in biophysics at Johns Hopkins because the programme was one of the few to explore the growing field's variety of research areas. He excelled, creating a method to simplify the process of determining a crystal's structure; postdocs followed at Johns Hopkins and the Max Planck Institute for Biochemistry in Martinsried, Germany. 'Lattman angles' are now used in crystallography computer programs around the world to display how molecules are oriented in space.

Lattman was later recruited to chair his former department, Johns Hopkins' Krieger School of Arts and Science, where he became his PhD adviser's boss. A series of retirements coupled with administrative mis-steps had left the biophysics department in a slump. Lattman hired new, diverse faculty to restore his graduate department's broad capabilities in biophysics.

Lattman now plans to strengthen Hauptman-Woodward by adding new investigators in areas such as computational biology and single-molecule diffraction, forging new collaborations with University of Buffalo researchers and raising endowments to guard against a nationwide funding crunch affecting many institutions. "Ed welcomes new ideas," says Johns Hopkins biophysics colleague George Rose. "And, importantly, he can instinctively spot the difference between a fresh approach and a quixotic gesture."

Virginia Gewin

## NETWORKS & SUPPORT

### Crossing borders

Extending an invitation can do more than bring in a guest — it can build a beneficial relationship. In 2005, the Feinstein Institute for Medical Research in New York named cytokine specialist Jan Andersson, head of infectious diseases at the Karolinska Institute of Stockholm, as its first invited lecturer. Andersson reciprocated by suggesting that the two institutes combine educational forces.

This month, they announced a joint training programme that will let Karolinska graduates do postdoctoral research in New York while Feinstein scientists have access to Karolinska labs. Both institutions have records of collaboration and mentorship that will be strengthened by the relationship, says Feinstein director Kevin Tracey. He and Andersson, for example, have worked together on the role of cytokines in inflammation.

The Feinstein Institute's research enterprise, with 750 employees doing research in 45 programmes, is about a third the size of the Karolinska's and lacks some of its infrastructure, such as mass spectrometry and proteomics facilities. However, it has strengths in inflammation research, neuroscience and neuropsychology that the Karolinska wants to tap into. "The Karolinska is much larger and has more departments, people and

programmes," Tracey says.

Meanwhile, the Karolinska wants to augment its translational and clinical research. The Feinstein can help, as it is affiliated with the North Shore–Long Island Jewish Health System, which boasts 38,000 patients in 15 hospitals; 120,000 patients have been enrolled in clinical trials since 1998. The Feinstein serves the greater New York area, with 8 million people, close to Sweden's total population.

The Karolinska's medical and graduate training is strong, says Tracey, "but they were interested in having help with their postdoctoral training."

Both institutions offer high-quality mentoring. The Karolinska Institute launched its junior faculty programme in 2004, helping graduates make the transition to independent researcher (see *Nature* **427**, 470; 2004). The Feinstein Institute specializes in making doctors who have practised medicine into researchers. And most do continue in research, unlike graduates from many other such programmes.

Each institution will host up to six researchers. Those interested in working in the other institute's lab must apply for a place and be reviewed by a joint admission committee, including scientists from both institutes.

Paul Smaglik

#### POSTDOC JOURNAL

### Meeting our targets

Non-scientists think research is not a 'real job' because scientists don't have deadlines or discrete targets. On the contrary, I say, we are accountable to those who fund us. Investors want financial returns, taxpayers demand medical advances. We are judged by the number and impact factor of our publications. I have friends who have actually calculated the average impact factor of papers published at their institution, hoping to gauge their own competitiveness.

At the Biopolis, we are subject to annual evaluations in which our productivity is rated on a scale of 1 to 5. These determine our bonus and possibly our career prospects. You don't get points for effort: my decision to study a novel protein has proved unwise because I cannot use established reagents and protocols to churn out data and papers. I can ask many questions about a protein of unknown function. But as I was specifically instructed to "focus on publishing my work as it is completed", the key question driving my research must be: what is the minimum amount of data that can coalesce into a paper?

This practical approach is difficult to reconcile with the risk associated with novel or creative projects. It is, however, necessary if I want to stay employed in a world dictated by 'key performance indicators'. We are subject to constant selection pressure. And, to paraphrase Darwin, those who can adapt, survive.

Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.





# THE NATIONAL INSTITUTES OF HEALTH WWW.NIH.GOV

## Fellowships in Cancer Epidemiology and Genetics, National Cancer Institute

### Discovering the Causes of Cancer and the Means of its Prevention

The Division of Cancer Epidemiology and Genetics (DCEG), an intramural research program of the National Cancer Institute (NCI), National Institutes of Health (NIH), conducts population and multidisciplinary research to discover the genetic and environmental determinants of cancer and new approaches to cancer prevention.

### Fellowship Opportunities

DCEG provides research and training resources and senior faculty for highly distinctive postdoctoral and predoctoral fellowships, including those of a multidisciplinary nature. DCEG fellows design, carry out, analyze, and publish research studies related to the etiology of cancer in human populations, with opportunities to work on large prospective cohort studies and international collaborations. Research topics encompass the full range of cancer risk factors as well as biostatistical approaches. Emphasis is placed on the inclusion of genomic and other emerging technologies into epidemiologic study designs to uncover genetic/environmental determinants and pathways of cancer induction and progression.

### A Multi-Faceted Research Program

Through its programs in cancer epidemiology, genetics, statistics, and related areas, DCEG:

- (1) Conducts broad-based, high-quality, high-impact research into cancer etiology and prevention;
- (2) Maintains a national and international perspective, giving priority to emergent issues identified through clinical, laboratory, and epidemiologic observations, as well as to public health concerns identified by the Institute, Congress, regulatory agencies, and other appropriate bodies;
- (3) Develops infrastructures, resources, and strategic partnerships in molecular epidemiology, including genome-wide association studies involving intramural/extramural consortia; and

- (4) Trains the next generation of scientists in cancer epidemiology, genetics, and related fields.

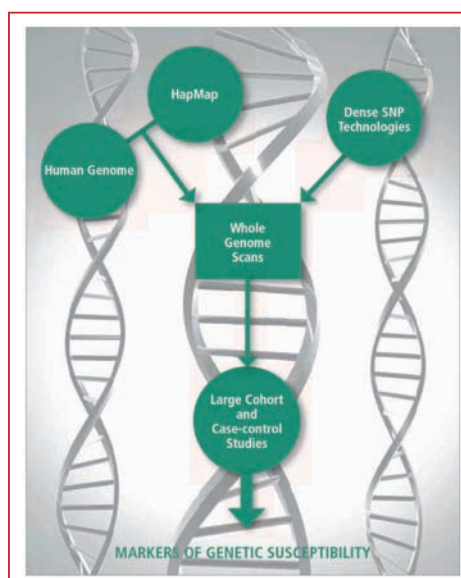
### Areas of Investigation

Major areas of investigation include:

- biostatistics and methodology
- clinical genetics
- descriptive epidemiology
- diet and nutrition
- hormonal factors
- exposure assessment
- gene-environment interactions
- hereditary syndromes
- molecular and genetic epidemiology
- second cancers
- occupational exposures
- pharmacoepidemiology
- pharmacogenetics
- radiation
- risk assessment

To learn more about the Division visit our web site on: <http://dceg.cancer.gov>. To inquire about fellowships contact: the DCEG Office of Education by email: [ncidceged-r@mail.nih.gov](mailto:ncidceged-r@mail.nih.gov)

**Joseph F. Fraumeni, Jr., M.D.**  
**Director, Division of Cancer Epidemiology and Genetics, NCI, NIH**





## Center for Cancer Research

### DIRECTOR, CANCER STEM CELL and DEVELOPMENTAL BIOLOGY PROGRAM

Application Deadline: May 30, 2008

The National Cancer Institute (NCI) is seeking an outstanding scientist to serve as the first *Director of the Cancer Stem Cell and Developmental Biology Program* in the Center for Cancer Research (CCR). This is an exciting opportunity to establish a nationally recognized program in stem cell and developmental biology research. The Director will lead a collaborative community of investigators who are committed to advancing this field with the goal of fostering novel therapeutic approaches that target cancer stem cells. The program will be supported by a wide array of intellectual and technological resources, including dedicated, high quality technology cores relevant to stem cell biology, such as imaging/visualization (both in vitro and in vivo), in vitro stem cell assays, transgenic models, genetics, gene/drug delivery, FACS, proteomic analysis (including high-throughput mass spectrometry), and human genetics/bioinformatics. In addition, the cancer stem cell and developmental biology initiative will be supported with stable financial resources and a training program in cancer stem cell and developmental biology for talented post-doctoral fellows.

The National Cancer Institute is part of the National Institutes of Health (NIH) in the Department of Health and Human Services (DHHS), a federal government agency. The Center for Cancer Research is the largest component of the intramural biomedical research effort at NIH. The research environment is collaborative and highly conducive to advancing translational research, emphasizing multidisciplinary and interdisciplinary team science. For more information go to: <http://ccr.cancer.gov/default.asp>.

Applicants must have a Ph.D., M.D., or M.D. /Ph.D. degrees and a documented history of research success. Currently tenured faculty or faculty eligible for tenure at NIH, with a demonstrated commitment to stem cell biology, are encouraged to apply. Salary will be commensurate with experience and accomplishments. Applications should include: A letter indicating the position of interest including a statement of research interests; a career synopsis; and a current *curriculum vitae* and complete bibliography.

Applications must be postmarked or received by email at [hooperl@mail.nih.gov](mailto:hooperl@mail.nih.gov) by May 30, 2008. Send applications to: Douglas Lowy, M.D., Chair, Cancer Stem Cell and Developmental Biology Program Search Committee, c/o Laura Hooper, Executive Secretary, Center for Cancer Research, 31 Center Drive, 31/3A11, MSC 2440, Bethesda, MD. 20892-2440.



### IMMUNOEPIDEMIOLOGIST Tenure-Track/Tenured Investigator Position



The new Infections and Immunoepidemiology Branch (IIB, <http://www.dceg.cancer.gov/iib>) in the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), is recruiting for a tenure-track/tenured epidemiologist with experience and interest in the study of immunology and cancer.

Research at the IIB focuses on understanding the causes and mechanisms involved in the development of tumors linked to infectious agents and in understanding the mechanisms underlying successful immune responses as they relate to cancer risk and to successful responses to vaccination. Both infectious and host immunological factors are considered in the context of human epidemiological studies. Studies undertaken by the group are often large, international and include rich biological specimen components to permit in-depth evaluation of biological processes, including immunological and inflammatory mechanisms, involved in tumor development and vaccination responses using state-of-the-art techniques. The Branch has an active set of studies aimed at evaluating the role of numerous infectious agents (including EBV, HIV, HPV, HTLV-1, and KHSV) and tumor sites (including cervical cancer, gastric cancer, Hodgkin and non-Hodgkin lymphomas, lung cancer, nasopharyngeal carcinoma, and oral/oropharyngeal cancers). The Branch also houses the 7,500 woman community-based vaccine trial in Costa Rica designed to evaluate the efficacy and impact of a new HPV-16/18 virus-like particle vaccine, and an HPV immunology laboratory that complements field activities and allows interactive evaluation of hypotheses of immunological importance within our studies.

The successful candidate will receive research support from the intramural research program of NIH for conducting innovative studies aimed at elucidating inflammatory and immunological mechanisms of HPV-related or other cancers and/or vaccine responses. Applicants must have an M.D. and/or Ph.D. in epidemiology or a related field, and considerable post-doctoral experience in cancer epidemiology, molecular epidemiology, and/or tumor or vaccine immunology. A record of peer-reviewed publications in one of these fields is required. A demonstrated ability to lead complex epidemiologic investigations is highly desirable. The successful candidate should have strong analytical skills, a good understanding of biological/immunological processes, and a demonstrated ability to collaborate across disciplines. Strong oral and written communication skills are an important requirement. Applications will be evaluated on their ability to develop a creative, independent program of epidemiological research applicable to understanding immune determinants of successful vaccination against cancer causing infections and/or immunological determinants of HPV-related or other cancers; and to collaborate effectively in a multidisciplinary setting.

Interested individuals should send a cover letter, curriculum vitae, a brief summary of research experience, accomplishments and research interests and goals, copies of three publications or preprints, and three letters of reference to: **Ms. Judy Schwadron, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd. EPS/8073, Bethesda, MD 20892.**

Candidates should submit applications by **May 30, 2008**; however, the search will continue until a qualified candidate is found. Additional information about staff and ongoing research in the Division of Cancer Epidemiology and Genetics and in the Infections and Immunoepidemiology Branch is available at <http://www.dceg.cancer.gov>. Prospective applicants should send E-mail inquiries to **Allan Hildesheim, Ph.D., Branch Chief** ([hildesha@exchange.nih.gov](mailto:hildesha@exchange.nih.gov)). This position is subject to a background investigation. DHHS and NIH are Equal Opportunity Employers.



OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH

NIH131251R



# Build your own time machine

It's a unique set of experiences.

**Igor Teper**

Every time machine is a one-of-a-kind device, and building one requires perseverance, ingenuity and total concentration. You will also need a clean workspace, a flat surface and the usual assortment of tools. Some of the parts and materials described in these instructions may not be available, in which case you will have to improvise suitable replacements.

Start with the sunlight's reflection off the water on the afternoon your grandfather took you out on the lake in a rowboat, the last time you saw him alive. Stabilize it with the bobbing of the ducks on the boat's wake, the breeze caressing your arms and neck, causing goose bumps, and the insistent tug of the oars in your hands when he let you row.

Attach your little sister's tears after you plucked out the eyes of her favourite plush teddy bear and hid them under her pillow for her to find in the morning. It should fit loosely, leaving room for the ringing of your mother's screams in your ears as well as the dampness of the soil on your hands when you buried your favourite toy soldier's splintered remains after your sister smashed it in parent-sanctioned retribution.

Gently screw into place the trembling of your first lover's lips as they formed a smile when you told her that you loved only her and would forever; fasten onto it the set of her jaw when you said you'd become involved with someone else. You may need to lubricate the joint with the lies you told her in the final weeks of your relationship, each encompassing the last, like Russian nesting dolls, mixed in equal parts with her blind desperation to believe you.

Be very careful as you slip into the resulting groove the smell of the wilted get-well flowers you found on your doorstep upon returning from the ski trip with your friends after you'd told your parents you were too sick to come home for Thanksgiving. Softly tap in your mother's worried, desperate pleas in the half-dozen messages she'd left on your answering machine and connect them with the disappointment that quickly replaced anger on your father's face when you told him the truth.

Hook up the cold slickness of sweat in your palms as you signed the papers finalizing your promotion to manager. Plug in the warmth of your wife's cheek against yours when you embraced before you got into the corporate car to go to the airport for your first extended business trip, and bind



JACEY

it in place with the smell of her hair and the pressure of her fingers against your back.

Affix the constriction in your chest linked with the wobbliness in your knees as you stood on the boardwalk and watched your then best friend walk away after you told him it was nothing personal but you wouldn't be comfortable lending him money. The golden-pink of the sunset should slide smoothly into the smell of the ocean, the feel of salt coating your skin, and the grainy roughness of the wooden guardrail in your grip.

Install the numbness, spreading from your fingertips to your hand and up your arm, after you hit your wife on the side of her head when, during an argument that started over nothing and touched on everything, she called you a series of names no wife should call her husband. Fill in any gaps with the sound of her shocked gasp and the heartbreaking silence that followed.

Snap in the sound of the door slamming when your son stormed out of your house after you explained to him, in increasingly forceful terms, the reasons that you'd be more than happy to help pay for law school but wouldn't give a penny for a creative-

writing programme. Latch onto it the flash of defiance in his eyes, the knot in the pit of your stomach and the scent of daffodils that drifted in from your front yard through the open door.

Surmount the whole thing with the tightness in the corners of your mother's mouth when you told her you would not be able to take care of her the way she needed if she came to live with you. Prop it in place with the trembling of her hands in yours, and the musty, medicinal smell of her apartment, which coated the inside of your nostrils, dried out your throat and stung your eyes.

If you've followed these instructions, and you're very lucky, you should have in your possession your very own, fully functioning, custom-made time machine, ready to take you to the moment of your choice in the blink of an eye.

If you're even luckier, you will leave it in the corner of the garage, next to the old lawnmower you've never got around to fixing, under a worn tarp, collecting dust. ■ **Igor Teper teaches old atoms new tricks at temperatures near absolute zero. He also writes stories, occasionally.**